

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**FACULTAD DE ECONOMIA Y PLANIFICACIÓN**



**“CLASIFICACIÓN DE RENDIMIENTO EN LA PRUEBA DE  
ADMISIÓN DE LA UNALM UTILIZANDO ANÁLISIS  
DISCRIMINANTE LINEAL Y MÁQUINA DE SOPORTE  
VECTORIAL”**

**Presentado por:**

**Fredy Vivanco Huaytara**

**TESIS PARA OPTAR EL TÍTULO DE  
INGENIERO ESTADÍSTICO E INFORMÁTICO**

**Lima – Perú**

**2018**

## **DEDICATORIA**

Este trabajo va dedicado a los dos pilares más importantes de mi vida, mis padres Isaac Vivanco Soto y Cristina Huaytara Huamán que con su infinito amor, esfuerzo y apoyo pudieron educarme y guiarme con el único fin de alcanzar mis sueños.

A mi hermano Dario Vivanco, por ser mi ejemplo de superación y enseñarme que todo esfuerzo en la vida es recompensado con al menos una sonrisa y a Nancy Vivanco por su paciencia.

A Katy Quispe Guizado por su ejemplo de amor y superación junto a nuestro más grande tesoro: Luciana Vivanco Quispe

A mi asesor Ing. Fernando Rosas Villena por su guía incondicional en el éxito de este trabajo, al ing. Jesús Salinas Flores por su valioso aporte en este trabajo y ser mi referente académico en la universidad y a Merajad Dumet Montoya por su gran ayuda

A mis amigos de la universidad, en especial a Paul Loarte Laos por ser su ejemplo de disciplina y sus enseñanzas.

A mi directora Sandra Romero Herrera por ser un referente de liderazgo y calidad humana.

A Cesar Ruiz Bancayan por la oportunidad de abrirme camino en este maravilloso mundo de la docencia.

# I. ÍNDICE GENERAL

II. RESUMEN.....	9
III. INTRODUCCIÓN.....	10
IV. REVISIÓN DE LA LITERATURA.....	13
4.1 Centro de Estudios Preuniversitarios.....	13
4.1.1 Reseña histórica.....	13
4.1.2 Formación académica.....	13
4.1.3 Ingreso a la Universidad Nacional Agraria la Molina.....	14
4.2 Técnicas de clasificación.....	15
4.2.1 Técnicas estadísticas de clasificación.....	16
4.2.2 Técnicas no estadísticas de clasificación.....	17
4.3 Análisis Discriminante Lineal.....	19
4.3.1 Definición.....	19
4.3.2 Objetivos.....	20
4.3.3 Función discriminante lineal.....	21
4.3.4 Pasos de la metodología del Análisis Discriminante Lineal de Fisher.....	23
Paso 1: Objetivos.....	24
Paso 2: Diseño de la investigación.....	24
Paso3: Supuestos.....	25
Paso 4: Estimación de la Función Discriminante Lineal de Fisher.....	26
Paso 5: Interpretación de los resultados.....	27
Paso 6: Validación de los Resultados.....	28
4.4 Máquina de Soporte Vectorial.....	28
4.4.1 Definición.....	28
4.4.2 Objetivos de la Máquina de Soporte Vectorial.....	29
4.4.3 Funciones kernel.....	30
4.4.4 Clasificador máximo margen.....	32
Caso: Poblaciones Separable Linealmente.....	33
Caso: Poblaciones No Separable Linealmente.....	36

4.4.5 Construcción de la Máquina de Soporte Vectorial.....	40
4.4.6 Pasos de la metodología de la Máquina de Soporte Vectorial .....	43
Paso 1: Objetivos .....	43
Paso 2: Diseño de la investigación .....	44
Paso 3: Supuestos .....	45
Paso 4: Estimación de la Máquina de Soporte Vectorial.....	46
Paso 5: Interpretación de los resultados .....	48
Paso 6: Validación de los resultados .....	48
4.5. Evaluación de la clasificación .....	49
V. MATERIALES Y MÉTODOS .....	52
5.1 Metodología de la investigación.....	52
5.1.1 Tipo de investigación .....	52
5.1.2 Formulación de hipótesis.....	52
5.1.3 Identificación de variables.....	53
5.1.3 Definición operacional de las variables.....	54
5.1.4 Diseño de la investigación.....	54
5.1.5 Población y muestra .....	55
5.1.6 Instrumento de colecta de datos .....	55
5.2 Metodología aplicada .....	55
VI. RESULTADOS Y DISCUSION .....	56
6.1 Resultados aplicación 1 .....	56
6.1.1 Exploración de variables: .....	56
6.1.2 Aplicación de las técnicas de clasificación .....	59
6.1.3 Evaluación de las técnicas de clasificación .....	65
6.2 Resultados aplicación 2 .....	69
6.2.1 Exploración de variables: .....	69
6.2.2 Aplicación de las técnicas de clasificación .....	72
6.2.3 Evaluación de las técnicas de clasificación.....	77
6.3. Comparación de Resultados: .....	82
VII. CONCLUSIONES .....	83

VIII. RECOMENDACIONES .....	85
IV REFERENCIAS BIBLIOGRÁFICA .....	86
X. ANEXOS: .....	91

## ÍNDICE DE CUADROS

Cuadro N° 1: Partición data total.....	45
Cuadro N° 2 Elementos de la Matriz confusión .....	49
Cuadro N° 3: Resultados Descriptivos .....	57
Cuadro N° 4: Factor de Inflación de la Varianza .....	61
Cuadro N° 5: Anova Univariado .....	62
Cuadro N° 6: Significancia de la Función Discriminante .....	62
Cuadro N° 7: Capacidad Discriminate de la Función Discriminante.....	63
Cuadro N° 8: Regla de Clasificación Kernel Lineal .....	63
Cuadro N° 9: Regla de Clasificación Kernel Radial .....	64
Cuadro N° 10: Validación de Resultados .....	64
Cuadro N° 11: Validación de Resultados .....	65
Cuadro N° 12: Validación de Resultados .....	65
Cuadro N° 13: Evaluación de la Clasificación .....	65
Cuadro N° 14: Evaluación de la Clasificación .....	66
Cuadro N° 15: Evaluación Clasificación Kernel Lineal.....	67
Cuadro N° 16: Evaluación Clasificación Kernel Radial .....	68
Cuadro N° 17: Resultados Descriptivos .....	70
Cuadro N° 18: Factor de Inflación de la Varianza .....	73
Cuadro N° 19: Anova Univariado .....	74
Cuadro N° 20: Significancia de la Función Discriminante .....	75
Cuadro N° 21: Capacidad Discriminate de la Función .....	75
Cuadro N° 22: Regla de Clasificación Kernel Lineal .....	76
Cuadro N° 23: Regla de Clasificación Kernel Radial .....	76
Cuadro N° 24: Validación de Resultados .....	77
Cuadro N° 25: Validación de Resultados .....	77

Cuadro N° 26: Validación de Resultados .....	77
Cuadro N° 27: Evaluación de la Clasificación .....	78
Cuadro N° 28: Evaluación de la Clasificación .....	78
Cuadro N° 29: Evaluación Clasificación Kernel Lineal.....	79
Cuadro N° 30: Evaluación Clasificación Kernel Radial .....	80
Cuadro N° 31: Comparación de las clasificaciones .....	82

## INDICE DE GRÁFICOS

Gráfico N° 1: Noción Función Kernel.....	31
Gráfico N° 3: Poblaciones Separable Linealmente .....	33
Gráfico N° 3: Poblaciones no separables Linealmente .....	37
Gráfico N° 4: Curva ROC .....	51
Gráfico N° 5: Gráfico Descriptivo Ingresantes y no Ingresantes .....	57
Gráfico N° 6: Gráfico Descriptivo Según condición de Ingreso .....	58
Gráfico N° 7: Área Bajo la Curva ROC .....	66
Gráfico N° 8: Área bajo la Curva ROC Kernel Lineal.....	67
Gráfico N° 9: Área bajo la Curva ROC Kernel Radial .....	68
Gráfico N° 10: Gráfico Descriptivo Ingresantes y no Ingresantes .....	70
Gráfico N° 11: Gráfico Descriptivo Según condición de Ingreso .....	71
Gráfico N° 12: Área bajo la Curva ROC.....	79
Gráfico N° 13: Área bajo la Curva ROC Kernel Lineal.....	80
Gráfico N° 14: Área bajo la Curva ROC Kernel Radial .....	81



## **II. RESUMEN**

En objetivo general de la investigación es la identificación entre las técnicas del Análisis Discriminante Lineal de Fisher y Máquina de Soporte Vectorial la que presenta mejores indicadores de clasificación del rendimiento de los postulantes en la prueba de admisión 2015-II en la Universidad Nacional Agraria La Molina (UNALM). Ambas técnicas estadísticas se aplicaron en dos oportunidades, en la primera se evaluó el resultado en la prueba admisión de la UNALM de los postulantes que no se prepararon en el CEP-UNALM y que ingresaron o no a la universidad (aplicación 1) y en la segunda se evaluó resultado en la prueba admisión de la UNALM de los postulantes que si se prepararon en el CEP-UNALM y que ingresaron o no a la universidad (aplicación 2). Los resultados muestran que la técnica Máquina de Soporte Vectorial presenta mejores indicadores de clasificación que el Análisis Discriminante Lineal de Fisher.

## **II. SUMMARY**

The aim of this research is the identification between the techniques of Fisher's Linear Discriminant Analysis and Vector Support Machine which presents better indicators of classification performance from the applicants in the admission test 2015-II Universidad Nacional Agraria La Molina (UNALM). Both statistical techniques were applied on two occasions; the first one evaluated the results of the UNALM admission test for applicants who were not prepared at CEP-UNALM and who entered university (application 1) and at the university. Second, the results were evaluated in the UNALM admission test of the applicants, if they were prepared in the CEP-UNALM and who entered university or not (application 2). The results show that the Vector Support Machine technique presents better classification indicators than the Fisher Linear Discriminant Analysis

### III. INTRODUCCIÓN

Este nuevo milenio ha traído no solamente la masificación y fácil acceso a la información y el conocimiento, sino también la creación de nuevas universidades en el país cuyas puertas de ingreso se abren a los postulantes a través de un examen de admisión. Gran parte de ellas, cuentan con academias o centros de estudios preuniversitarios, cuya finalidad es preparar a los postulantes para sortear con éxito el examen de admisión, reducir en algunos casos la brecha entre la formación escolar y la universitaria y familiarizar a los ingresantes con las exigencias académicas de la universidad.

La Universidad Nacional Agraria La Molina (UNALM) fue la primera institución de educación superior del país en crear en el año 1947, una academia de preparación preuniversitaria con el nombre de Centro Pre Universitario y que en la actualidad lleva el nombre de Centro de Estudios Preuniversitarios de la Universidad Nacional Agraria La Molina (CEP-UNALM).

El plan de estudios del CEP-UNALM contempla la formación de sus estudiantes en las siguientes áreas del conocimiento: Matemáticas, Física, Química, Biología, Razonamiento Matemático y Razonamiento Verbal. El área de matemáticas está conformada por los cursos de Aritmética, Algebra, Geometría y Trigonometría. Todas estas materias son objeto de evaluación en la prueba de admisión de la UNALM. Por esta razón uno de los principales objetivos del CEP-UNALM, es que sus alumnos ingresen a la universidad con un rendimiento diferenciado en la prueba de admisión respecto a los postulantes que no se prepararon en ella, demostrando a la sociedad la calidad de los servicios académicos que brinda.

En Estadística, el término “discriminar” o “clasificar” se utiliza para caracterizar las clases o categorías que adopta una variable dependiente sobre la base del comportamiento de un conjunto de variables independientes. Existen diversas técnicas cuyo objetivo es la identificación de una regla de clasificación que permita predecir la pertenencia de los nuevos elementos a uno de los grupos de clasificación. Las técnicas de clasificación se pueden clasificar en supervisadas y no supervisadas. Las técnicas

supervisadas se caracterizan porque las clases o categorías de la variable dependiente están previamente definidas, a diferencia de las técnicas no supervisadas donde no se definen previamente.

El Análisis Discriminante Lineal de Fisher es una de estas técnicas de clasificación supervisada. Su naturaleza es paramétrica y su objetivo es encontrar una combinación lineal de las variables independientes que mejor permitan diferenciar a tales grupos por medio de una función discriminante lineal. Esta función está basada en la descomposición de la variabilidad total de la muestra. La función se fundamenta cuando el ratio variabilidad dentro de los grupos y entre los grupos se maximiza.

La Máquina Soporte Vectorial (SVM) es también una técnica de clasificación supervisada, perteneciente a la Inteligencia Artificial aplicada o Aprendizaje Automático que según Ali(2006) y Shigeo (2005) presenta mejor rendimiento en comparación con muchas técnicas debido a su naturaleza no paramétrica y por ser de tipo computacional o algorítmica. Esta técnica es no lineal y se basa en clasificadores de margen máximo, donde el objetivo de un clasificador de margen máximo es encontrar una superficie de decisión/hiperplano o función ( $\mathbf{w}^T \mathbf{x} + b = 0$ ) que sea equidistante a las fronteras de los grupos justo en aquel lugar en el que los grupos se encuentran más próximas y maximizar el margen existente entre tales grupos.

Estas dos técnicas de clasificación tienen el mismo objetivo, pero estableciendo procedimientos diferentes en sus metodologías y en la forma de abordar el problema de clasificación. El Análisis Discriminante Lineal de Fisher utiliza la función Discriminante Lineal para maximizar la separabilidad entre los grupos y reducir al mínimo la variabilidad dentro del grupo. La técnica Máquina Vectorial de Soporte lo hace a través de la superficie de decisión que generalmente mediante la aplicación de una función Kernel determinada maximiza el margen existente entre los grupos. Es debido a esta propiedad, que algunos autores como Hsu(2003) afirman que la técnica Máquina de Soporte Vectorial tendría mejores indicadores de clasificación que la técnica Análisis Discriminante Lineal de Fisher.

La clasificación en clases o grupos se logra a través de un modelo o regla generada que sigue una estructura según la técnica empleada, ya sea en términos de una Análisis Discriminante de Fisher o mediante La Máquina de Soporte Vectorial, lo mismo que se obtiene con una muestra especificada, llamada muestra de entrenamiento el cual es una porción del conjunto de datos total.

Para medir el poder clasificatorio y/o clasificar futuras observaciones con dicha técnica se emplea otra muestra especificada proveniente de la muestra inicial total, esta muestra se llama de validación y de prueba respectivamente.

En los últimos tiempos se han venido desarrollando investigaciones en donde se comparan técnicas de clasificación clásicas frente a las de reciente desarrollo, dentro de ellas están las comparaciones del Análisis Discriminante Lineal de Fisher y la Máquina de Soporte Vectorial respectivamente. Los resultados en términos de su capacidad clasificatoria son divergentes. Un ejemplo de ello, es el estudio realizado por Lee et al (2005) en donde el Análisis Discriminante Lineal tiene un mejor performance de clasificación que la técnica Máquina de Soporte Vectorial a diferencia del obtenido por Pitarque et al (1998). Algunos autores consideran que estas discrepancias empíricas son más producto de una incorrecta aplicación de las técnicas que de su poder clasificatorio.

Por esta razón, el objetivo principal de la presente investigación es medir y comparar la capacidad de clasificación de las técnicas Análisis Discriminante Lineal de Fisher y Máquina Vectorial de Soporte. Para ello, se utiliza la matriz de confusión con sus indicadores; especificidad, sensibilidad, exactitud, valor predicho (positivo y negativo), así como el área bajo la curva ROC.

El estudio considera a los postulantes a la UNALM de la modalidad del Concurso de Admisión Ordinario 2015-II, los mismos que son identificados y clasificados según su preparación o no CEP-UNALM y su condición de ingresante o no a la universidad.

## **IV. REVISIÓN DE LA LITERATURA**

### **4.1 Centro de Estudios Preuniversitarios**

#### **4.1.1 Reseña histórica**

Los orígenes de la Universidad Nacional Agraria La Molina se remontan a 22 de Julio de 1902 cuando la misión belga planificó y organizó en el fundo de Santa Beatriz la Escuela Nacional de Agricultura y Veterinaria (ENAV), Años más tarde la Escuela es cambiado de sede al distrito de la Molina donde permanece hasta el día de hoy. Su consolidación institucional vendría años más tarde cuando en 1960 La Ley Universitaria N° 13417 reconoce a la Escuela Nacional de Agricultura su rango universitario con la denominación Universidad Nacional Agraria La Molina.

En 1947 bajo la dirección del Dr. J. Alberto León se la crea la sección preparatoria con el objetivo de diferenciarla de una academia de preparación más y con el propósito de integrar la organización de la Escuela Nacional de Agricultura y Veterinaria de ese entonces. El 4 de Setiembre de 1979, la Universidad Nacional Agraria La Molina, mediante la Resolución N° 35920 / UNA aprueba el Reglamento de la sección preparatoria, durante el rectorado del Ing. Mario Zapata Tejerina, finalmente en febrero de 2004, se le denomina Centro de Estudios Pre Universitarios (CEP-UNALM), mediante Resolución N° 004-2004/UNALM-AU.

#### **4.1.2 Formación académica**

El CEP-UNALM es la decana de los centros pre universitarios en el Perú (1947-2017) alberga un promedio de 1,800 alumnos por cada ciclo regular donde por reglamento se ofrece una cobertura del 30% de las vacantes de ingreso.

El servicio brindado parte desde los aspectos formativos del alumnos que a través de un departamento psicopedagógico ofrece el soporte de asesorías y una orientación permanente sin descuidar el reforzamiento de las técnicas de estudio necesarias para el

logro de los mejores resultados, así mismo una educación y preparación que a través de grupos de estudio y seminarios conducidos por tutores especializados garantizan y aseguran las condiciones académicas más adecuadas para que los postulantes ingresen a la Universidad Nacional Agraria La Molina.

Las constantes evaluaciones se miden por el rendimiento de los cursos de ciencias básicas tales es el caso de Razonamiento Matemático, Razonamiento Verbal Matemática, Física, Química y Biología.

#### **4.1.3 Ingreso a la Universidad Nacional Agraria la Molina**

La UNALM cuenta en la actualidad con las siguientes modalidades de ingreso:

1. Traslados externos de otras universidades
2. Dos primeros puestos de colegios de educación secundaria
3. Colegio Mayor Secundario Presidente del Perú
4. Quinto año de secundaria
5. Graduados o titulados universitarios
6. Ley N° 27277 – Víctimas de Terrorismo
7. Ley N° 29973 – Personas con Discapacidad
8. Ley N° 28036 – Promoción y Desarrollo del Deporte
9. Ley N° 28592 – Plan Integral de Reparaciones
10. Bachillerato en convenio con UNALM
11. Convenio Andrés Bello
12. Becario extranjero
13. Diplomáticos y funcionarios internacionales extranjeros
14. Diplomáticos y funcionarios internacionales peruanos
15. Cónyuges e hijos de docentes ordinarios de carrera de la UNALM en comisión de servicios fuera del País
16. Convenio Centro Regional de Formación en Meteorología
17. Ingresantes del Centro de Estudios Preuniversitarios de la UNALM

La modalidad que alberga el mayor número de postulantes es la del Concurso Ordinario de Admisión, la misma que se realiza 2 veces por año y que registra en cada proceso al 51,6% del total de postulantes.

Los postulantes que concurren a este concurso ordinario de admisión proveniente de las diferentes partes del Perú y cuya formación académica provienen también de las diferentes casas de preparación existentes en el país. Se rinden una prueba que tiene un tiempo de aplicación de 3 horas y que consta de 100 preguntas con cinco alternativas cada una distribuidas en 6 materias de la siguiente forma: Razonamiento Matemático, Razonamiento Verbal Matemática, Física, Química y Biología, donde el rendimiento o resultado en estas materiales de cada postulante, discriminará a los Ingresantes o no Ingresantes a la Universidad Nacional Agraria La Molina.

#### **4.2 Técnicas de clasificación**

Un problema de clasificación consiste formalmente en asignar una etiqueta o clase a un objeto. Cada objeto se describe por un conjunto de variables que representan medidas u observaciones sobre el mismo, y se le asocia un vector en un espacio n-dimensional, donde cada dimensión representa una variable. Así, el objetivo es clasificar a un objeto

$X = [x_1, x_2, \dots, x_n] \in R^n$ , donde  $R^n$  se denomina espacio de características o espacio de variables. Cada objeto tiene asignado una etiqueta o clase llamada  $l(x)$ . Para un problema con C clases  $l(x)$  puede tomar c valores discretos diferentes. El objetivo entonces es encontrar una función clasificadora  $h$  tal que para cada  $X$  sea  $h(x)=l(x)$ .

El objetivo de las técnicas de clasificación es encontrar modelos o reglas (funciones o reglas) que describen y distinguen clases o conceptos para futuras predicciones. En la investigación científica muchos autores asocian a la clasificación con etiqueta de valores o clases como se ha descrito anteriormente, mientras que la predicción es asociada a valores continuos. Es decir, en el conjunto de entrenamiento la variable objetivo es una variable continua.

Finalmente, clasificación y predicción para varios autores vienen siendo lo mismo, aunque se pueden hacer la diferenciación según el tipo de variable.

El análisis predictivo es la mejor forma de pronosticar sucesos de incertidumbre cuando ya se ha establecido reglas empíricas o reglas estructuradas de clasificación, el mismo que dependerá a partir de las condiciones iniciales de los datos de entrada o datos históricos (número de observaciones y número de las variables a tratar). Las formas de predecir en función a tales reglas pueden variar, por ejemplo desde un juicio puro, que si bien es poco estructurado ya que se emplea poca cantidad de data histórica, pero que se basa en evaluaciones subjetivas y en base a experiencias pasadas. Muchas veces tiene una alta eficacia en casos particulares.

La era de la tecnología de la información ha ocasionado que sea necesario el desarrollo de sistemas o reglas de clasificación estructurados, el mismo que están basados en modelos, así Vera (2003) menciona “los modelos son una representación formal de un sistema real, con el que se pretende aumentar su comprensión, hacer predicciones y ayudar a su control”. Esto ha conllevado que un sin número de técnicas de clasificación hayan emergido en las diferentes disciplinas de la ciencia, lo mismo que ha hecho muy flexible su uso y aplicación en muchas áreas de la investigación que van desde la medicina, educación, riesgo crediticio, etc.

#### **4.2.1 Técnicas estadísticas de clasificación**

En el campo de la Estadística inicialmente los métodos de clasificación eran basados en los métodos discriminantes para un problema general de clasificación cimentado en algunos supuestos estadísticos.

Conforme la tecnología de la información ha avanzado se hizo necesario desarrollar técnicas cuya estructura no requiera de supuestos estadístico tan escritos como por ejemplo en el caso de la regresión logística. Pero a menudo los investigadores o usuarios de estas técnicas, llamados por algunos como métodos multivariantes, deben



seguir un esquema sistemático por ejemplo Hair et al (1999) propone un esquema de modelización o generación de la regla de clasificación, de varios métodos multivariantes, en seis pasos a) Objetivos del análisis de la técnica multivariante (b) Diseño de la investigación de la técnica multivariante (c) Evaluación de supuestos de la técnica multivariante (d) Estimación del modelo multivariante (e) Interpretación de resultados (f) Validación del modelo multivariante.

“La intención no es proporcionar un conjunto rígido de procedimientos a seguir, sino más bien ofrecer una serie de pautas que facilitan la aproximación al diseño de los modelos cada investigador debe desarrollar su propio criterio de “éxito o fracaso” en cada nivel.”

En conclusión, Las técnicas pertenecientes a esta disciplina son altamente estructuradas, basadas en datos que han de cumplir ciertos supuestos y que constan de un esquema sistemático y flexible para su modelamiento y por ende las clasificaciones generadas por estas reglas son altamente confiables.

#### **4.2.2 Técnicas no estadísticas de clasificación**

Como se ha abordado anteriormente las formas de clasificación basadas en la generación de reglas pueden variar, desde un juicio puro que se basa en evaluaciones subjetivas y en base a experiencias pasadas.

La investigación científica ha hecho todo lo posible para que las Máquina sean un aliado muy importante en las diferentes actividades de la humanidad y el campo predicción o clasificación no ha sido la excepción. Es por este último que han emergido muchas disciplinas, así por ejemplo la inteligencia artificial aparece a mediados de los años cincuenta. Inicialmente con interrogantes como: ¿Pueden las máquinas aprender? Planteadas por Alan Turing a inicios del siglo 20 el cual es considerado padre de esta disciplina.

Investigadores como Alonso y Talbot (1992) definen a la Inteligencia artificial como “una disciplina científica que estudia la inteligencia humana utilizando el ordenador como principal herramienta de trabajo”.

La Inteligencia Artificial ha ido evolucionando conforme la tecnología de la información y la sociedad han ido desarrollándose. La consolidación de la inteligencia artificial en la década de los 70, se explica por la generación de reglas a partir de grandes volumen de datos a través de procedimientos iterativos y computacionales. Estas metodologías han ganado bastante terreno en las etapas finales del siglo pasado con la aparición de técnicas novedosas dentro de esta disciplina tales como Las Redes Neuronales o las Máquinas de Soporte Vectorial entre otras.

Se puede encontrar 2 áreas principales de la Inteligencia Artificial según Paniagua (2015), la inteligencia artificial robusta que en una perspectiva futurista las máquinas podrán ejecutar todas las habilidades cognitivas de los humanos y la Inteligencia artificial aplicada llamada también Aprendizaje Automático (Machine Learning) donde el aprendizaje (auto aprender de datos) se realiza a través de algoritmos, lo cual permite generar un Sistema Experto y ofrecer una eficaz clasificación basada en una regla o algoritmo.

La inteligencia artificial aplicada a través del Aprendizaje Automático (Machine Learning), se caracteriza por ofrecer buenos métodos de clasificación obtenida con soluciones iterativas, con un coste computacional aceptable y flexibles ala verificación los supuestos. Alonso y Talbot (1992) proponen “por qué no incorporar el conocimiento de los sistemas expertos en Estadística a los ordenadores, de forma que el usuario de los paquetes de análisis estadístico pueda beneficiarse del conocimiento de tales expertos de la Inteligencia artificial?”

Hair et al (1999) proponen como técnicas estadísticas de modelamiento al Análisis Discriminante y la Regresión Logística. Parr (2001) propone un proceso de modelamiento para un conjunto de técnicas clasificación dentro del aprendizaje automático en una serie de pasos que inicia en:

- (a) Selección de los datos
- (b) Preparación de los datos
- (c) Selección de las variables y transformación
- (d) proceso de modelamiento
- (e) Validación del modelo.

Si bien lo anterior parece tener una nomenclatura diferente al de Hair et al (1999) el objetivo de su aplicación viene a ser lo mismo en términos de clasificación para ambas disciplinas.

Autores como Jammes et al (2014) han unificado las diversas técnicas de aprendizaje automático y las técnicas estadística en un campo llamado Aprendizaje Estadístico, tal es el caso de la Máquina de Soporte Vectorial y el Análisis Discriminante de Fisher respectivamente, donde el procedimiento de modelamiento en esencia es similar para dichas técnicas. Ambos son metidos a un conjunto de datos llamado muestra de entrenamiento, el cual un algoritmo realiza el proceso de aprendizaje o entrenamiento y una segunda muestra que se ha mantenido al margen durante el proceso de aprendizaje llamado conjunto de datos o muestra prueba que será de utilidad luego para realizar o evaluar el poder de clasificación de la técnica que se ha entrenado.

### **4.3 Análisis Discriminante Lineal**

#### **4.3.1 Definición**

Pérez (2007) señala que: “El análisis discriminante es una técnica estadística utilizada para analizar la relación entre una variable dependiente (o endógena) no métrica (categórica) y varias variables independientes (o exógenas) métricas”.

Hair et al (1999) mencionan que: “El análisis discriminante es la técnica estadística apropiada para contrastar la hipótesis de que las medias de los grupos de un conjunto de variables independientes para dos o más grupos son iguales”.

Levy y Varela (2003) indican que es una técnica de clasificación y asignación de un elemento a un grupo del que se conocen algunos determinados atributos. El perfil de clasificación de los distintos elementos de una población o grupo se realiza a través de funciones de carácter lineal o cuadrático, que combinan las variables más influyentes en la pertenencia de los elementos a los grupos predefinidos, permitiendo interpretar las razones de su agrupamiento y pronosticar su comportamiento en relación con su pertenencia o no a los grupos objeto de análisis.

#### **4.3.2 Objetivos**

Para Hair et al (1999), el análisis discriminante puede abordar los siguientes objetivos de investigación: (a) establecer los procedimientos para clasificar elementos dentro de grupos, en base a sus puntuaciones sobre un conjunto de variables independientes, (b) determinar cuál de las variables independientes cuantifica mejor las diferencias en los perfiles de las puntuaciones medias de dos o más grupos, (c) determinar si existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre un conjunto de variables de dos o más grupos definidos a priori.

Pérez (2007) indica que: “El objetivo esencial del Análisis Discriminante es utilizar los valores conocidos de las variables independientes para predecir con qué categoría de la variable dependiente se corresponden”.

De estas definiciones de objetivos se puede afirmar que el objetivo principal del análisis discriminante es la descripción de diferencia entre grupos y la predicción de pertenencia a los grupos en función de ciertas variables o características conocidas para los sujetos por las variables independientes métricas.

### 4.3.3 Función discriminante lineal

Hair et al (1999) definen la función discriminante lineal para el análisis discriminante como una ecuación que tiene la siguiente forma:

$$Z_p = a + W_1 X_{1k} + W_2 X_{2k} + W_3 X_{3k} + \dots + W_n X_{nk} \quad (4.1)$$

Donde:

$Z_{jk}$  : Puntuación Z discriminante de la función discriminante j para el objeto k

$a$  : Constante

$W_i$  : Ponderación discriminante para la variable independiente i

$X_{ik}$  : Variable independiente i para el objeto k

La función discriminante propuesta por Fisher (1936) es la siguiente:

$$D_i = U_1 X_{i1j} + U_2 X_{i2j} + \dots + U_{1n} X_{inj} \quad (4.2)$$

Donde:

$i = 1, 2, \dots, k$  (grupos; en el caso de dos grupos,  $k = 2$ )

$j = 1, 2, \dots, n$  (observaciones)

La variable D es la variable a determinar con mayor poder discriminante. El criterio para obtener el eje discriminante D consiste en maximizar el ratio:

$$\frac{\text{Variabilidad entre Grupos}}{\text{Variabilidad Intra Grupos}} \quad (4.3)$$

Para Uriel et al (2009) Con este criterio se trata de determinar el eje discriminante de forma que las distribuciones proyectadas sobre el mismo estén lo más separado posible entre sí (mayor variabilidad entre grupos) y, al mismo tiempo, que cada una de las distribuciones este lo menos disperso (Menor variabilidad dentro de los grupos).

Las puntuaciones discriminantes se corresponden con los valores obtenidos al proyectar cada punto del espacio k-dimensional de las variables originales sobre el eje discriminante

### Notación Matricial

$$\mathbf{d} = \mathbf{XU} \quad (4.4)$$

La variabilidad de la función discriminante (suma de cuadrados de las desviaciones de las variables discriminantes con respecto a su media) se expresa:

$$\mathbf{d'd} = \mathbf{u'X'Xu} \quad (4.5)$$

$\mathbf{X'X}$  Es una matriz simétrica que se expresa las desviaciones cuadráticas con respecto a la media de las variables (suma de cuadrados total). Se puede descomponer en suma de cuadrados entre grupo F y suma de cuadrados intra grupo V.:  $\mathbf{X'X} = F + V$

Los ejes discriminantes vienen dados por los vectores propios asociados a los valores propios de la matriz  $V^{-1}F$  ordenados de mayor a menor.

### Clasificación con dos grupos

Los centros de gravedad o centroides son los estadísticos básicos que resumen la información sobre los grupos. Los centroides en caso de 2 grupos serán lo siguiente:

$$\bar{x}_I = \begin{bmatrix} \bar{x}_{1,I} \\ \bar{x}_{2,I} \\ \bar{x}_{k,I} \end{bmatrix} \quad \bar{x}_{II} = \begin{bmatrix} \bar{x}_{1,II} \\ \bar{x}_{2,II} \\ \bar{x}_{k,II} \end{bmatrix}$$

Con lo que, para los grupos I y II se obtiene:

$$\begin{aligned}
 D_I &= U_1 X_{1,I} + U_2 X_{2,I} + \dots + U_K X_{K,I} \\
 D_{II} &= U_1 X_{1,II} + U_2 X_{2,II} + \dots + U_K X_{K,II}
 \end{aligned}
 \tag{4.6}$$

El punto de corte discriminante C se calcula mediante el promedio:

$$C = \frac{\bar{D}_I + \bar{D}_{II}}{2}$$

El criterio para clasificar el individuo i:

*Si:  $D_i < C$ , se clasifica al individuo i en el grupo I*

*Si:  $D_i > C$ , se clasifica al individuo i en el grupo II*

En general, cuando se aplica el análisis discriminante se le resta el valor de C a la función discriminante (4.2). Por lo que, se clasifica en el grupo I si  $D-C > 0$ , y en el grupo II en otro caso.

#### **4.3.4 Pasos de la metodología del Análisis Discriminante Lineal de Fisher**

##### **Exploración de Variables**

Antes de emplear cualquier metodología de clasificación es importante realizar una revisión descriptiva tanto analítica como gráficamente, pues brindan una explicación del comportamiento de las variables previo a los resultados de cada técnica.

## **Paso 1: Objetivos**

Hair (1998) menciona que “el análisis discriminante es útil cuando el investigador está interesado en comprender las diferencias de los grupos o en clasificar correctamente objetos en grupos o clases... puede considerarse como una técnica predictiva analítica”

El análisis discriminante considera los siguientes objetivos de investigación:

1. Determinar si existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre un conjunto de variables de dos (o más) grupos definidos a priori.
2. Determinar cuál de las variables independientes cuantifica mejor las diferencias en los perfiles de las puntuaciones medias de dos o más grupos.
3. Establecer los procedimientos para clasificar objetos (individuos, empresas, productos, etc.).
4. Establecer el número y la composición de las dimensiones de la discriminación entre los grupos formados a partir del conjunto de variables independientes”

## **Paso 2: Diseño de la investigación**

En el segundo paso el investigador debe tomar decisiones sobre que variable deberá ser considerada como dependiente y que variables como independientes, sobre el tamaño de la muestra que va a considerar en el estudio y en la estrategia para dividir la muestra con fines de estimación de la función discriminante y otra con fines de validación.

En cuanto al tamaño de la muestra se recomienda utilizar un ratio de 20 observaciones por cada variable independiente que participe en el estudio, en los casos que sea imposible llegar a este ratio ideal se recomienda no usar un ratio que sea menor a 5. Además, de la decisión sobre el tamaño de muestra total el investigador debe tomar en cuenta que el tamaño de la muestra de cada grupo no debe ser menor que el número de variables predictoras y los tamaños relativos de cada grupo no varíen sustancialmente.



Para dividir la muestra se considera una muestra donde el 70% se considere para el desarrollo de la regla de clasificación y el 30% restante para la prueba de validación

### **Paso3: Supuestos**

Para la obtención de la función discriminante se requiere el cumplimiento ciertos supuestos, Para Hair et al ( 1999), un supuesto que siempre es necesario verificar es la de normalidad multivariada de las variables independientes, El incumplimiento del supuesto de normalidad multivariante puede hacer imprecisa la estimación de la función discriminante.

Los grupos deben proceder de la misma población y más en concreto es que estas matrices poblacionales correspondientes a cada grupo son iguales entre sí. Este supuesto alude a las matrices de varianzas-covarianza poblacionales debe ser la misma, Una prueba fiable es mediante el Estadístico M de Box, donde la hipótesis nula representa una igualdad de matrices-covarianzas poblacionales.

El incumplimiento de estos supuestos tanto de la normalidad Multivariada y matrices de varianzas-covarianza poblacionales iguales no presenta inconveniente con la consecución del análisis Discriminante, así para Muñoz (1998) indica: “No obstante, el no cumplimiento de estos supuestos no es problema para el análisis”.

Otro supuesto importante involucra a las variables redundantes, es decir diagnosticar y evitar la multicolinealidad entre las variables independientes ya que la presencia de correlación entre variables resta capacidad explicativa al total de variables. Una manera de cuantificar la multicolinealidad es mediante el factor de Inflación de la Varianza (FIV). El i-ésimo factor de inflación de varianza es la versión del coeficiente de correlación múltiple entre la variable independiente i-esima y el resto de variables independientes:

$$VIF_i = \frac{1}{1 - R^2}$$

Un factor de Inflación de la varianza cuyo valor supere 10, podría ser indicador potencial de problemas de multicolinealidad XinYao (2008)

Alternativamente varios autores entre ellos Montanero (2008) propone una alternativa diferente frente a las violaciones de los supuestos mencionados, como la aplicación del análisis discriminante cuadrático, en el caso en que las relaciones no sean lineales por ejemplo.

#### **Paso 4: Estimación de la Función Discriminante Lineal de Fisher**

Es preciso recalcar que Parr (2001) recomienda como paso previo a la estimación de la función discriminante, realizar la preparación de datos, que puede consistir en la transformación de los mismos frente a supuestos violados. De la misma forma realizar una selección variables independientes donde el objetivo es elegir las variables que sean relevantes, con ello lograr un rendimiento eficiente con el mínimo esfuerzo.

Se debe tener en cuenta que las variables independientes que poca importancia y redundantes existentes en una base de datos pueden tener un efecto negativo en los algoritmos de clasificación, según Aluja (2001) y Riquelme et al (2009) debemos tener presente: (a) Los atributos irrelevantes y los redundantes pueden “confundir” al algoritmo de aprendizaje, en el sentido de incorporar patrones que no guardan relación con el fenómeno a describir, por lo que en general, el clasificador obtenido es menos exacto que otro que aprenda sólo de datos relevantes. (b) La presencia de estos atributos provoca un mayor costo computacional para generar la función de clasificación ya que el tener más datos implica un mayor tiempo de procesamiento de ellos.

Para obtener la función discriminante se pueden utilizar dos métodos: el general (simultáneo) y el método por etapas. En el método general la función discriminante o funciones discriminantes se obtiene(n) con la participación simultánea de todas las variables independientes. En el método por etapas se incluyen las variables independientes en la función discriminante de una en una de acuerdo a su capacidad discriminatoria.

Luego de estimar la función discriminante se debe determinarse su significación, Para ello se cuenta con varios criterios, en el caso que se haya estimado la función discriminante por el método general se recomienda la medida del lambda de Wilks, la traza de Hotelling, el criterio de Pillai entre otros, y en el caso de una estimación por etapas se recomienda las medidas  $D^2$  de Mahalanobis y V de Rao.

El procedimiento de la  $D^2$  de Mahalanobis es el preferido cuando se está interesado en aprovechar al máximo la información disponible. Si el número de grupos es tres o más, entonces el investigador debe decidir no solamente si la discriminación entre el total de los grupos es estadísticamente significativa, sino también si cada una de las funciones discriminantes estimadas son estadísticamente significativas.

### **Paso 5: Interpretación de los resultados**

La interpretación de los resultados se realiza cuando la función discriminante es estadísticamente significativa y la precisión de la clasificación es aceptable a juicio del investigador. Bajo este procedimiento las funciones discriminantes son evaluadas para determinar la importancia relativa de cada variable independiente en la discriminación de los grupos. Existen tres métodos para determinar la importancia relativa: las ponderaciones discriminantes estandarizadas, las cargas discriminantes (correlaciones de estructura) y los valores parciales de la prueba F.

Las ponderaciones discriminantes o coeficientes discriminantes se usan para medir el grado de contribución que tiene una variable independiente con su función discriminante asociada. La contribución puede ser positiva o negativa dependiendo del signo que tenga el coeficiente dentro de la función.

Las cargas discriminantes denominadas correlaciones de estructura se usan alternativamente a los coeficientes debido a ciertas deficiencias de inestabilidad que puede presentar este último indicador. Las cargas discriminantes miden la correlación lineal simple entre cada variable independiente y la función discriminante. Así Hair (1999) detalla la importancia de las correlaciones de estructura: “Finalmente para

conocer cuáles son las variables que tienen mayor poder discriminante en orden a clasificar a un individuo en uno de los grupos. Una forma de medir ese poder discriminante es calculando el coeficiente de correlación entre cada una de las variables y la función discriminante”

En cuanto a los valores parciales de F, se utilizan cuando se generan las funciones discriminantes mediante el enfoque por etapas. En este caso los valores absolutos de F se clasifican en función a su tamaño, valores grandes indican capacidades discriminantes altas y valores pequeños capacidades discriminantes bajas.

## **Paso 6: Validación de los Resultados**

La validación de los resultados es la última etapa de su aplicación y permite asegurar que los resultados tengan validez externa e interna. Los procedimientos de mayor uso para verificar la validez de los resultados es la validación cruzada y el diseño de grupos.

En la validación cruzada se pueden dividir los grupos aleatoriamente en la muestra de análisis o ampliar dicha muestra para verificar su capacidad clasificatoria. El diseño de grupos es un procedimiento que consiste en perfilar los grupos en función a sus medias con el fin de asegurar su correspondencia con la fundamentación de su formulación.

## **4.4 Máquina de Soporte Vectorial**

### **4.4.1 Definición**

Para Vapnik (1995) Las Máquinas de Soporte Vectorial están basadas en el uso de los kernels y la construcción de un hiperplano de separación óptimo (Margen) en un contexto no paramétrico.

Tibshirani y Friedman (2008) definen a la Máquina de Soporte Vectorial como nuevas estructuras de aprendizaje basadas en la teoría estadística del aprendizaje basada en el principio de minimización estructural del riesgo, así mismo estos autores definen a esta

técnica como un problema donde la función discriminante es no lineal, donde una transformación del espacio de entrada (espacio original de datos) a otra dimensión superior (infinita), mediante la aplicación de un funciones no lineal llamada Kernel  $\Phi(x)$  deriva a que los datos sean linealmente separables así el problema puede ser resuelto mediante un hiperplano óptimo (de máximo margen) que delimite a las clases o grupos existentes.

García y Lozano (2006) definen de manera similar en el caso de una clase binaria: “ es encontrar el hiperplano separador óptimo (aquel que maximiza el margen) en el espacio de características el cual está relacionado con el espacio de entrada (espacio original de los datos) por medio de una transformación no lineal  $\Phi(x)$  (Kernel) de altas dimensiones que busca que los datos del conjunto de elementos etiquetados o que pertenecen a una misma clase queden al mismo lado, mientras se maximiza la distancia de dichas clases al hiperplano”

Para Hamel (2009) “Dicha transformación no lineal del espacio de entrada para que los datos se conviertan linealmente separables no implica un costo computacional en cambio es de vital importancia controlar los parámetros Máquina de Soporte Vectorial para lograr que esta técnica se ajuste bien a los datos”. Haciendo referencia a este último, para (Velásquez et al , 2010) la estimación de los parámetros de una Máquina de Soporte Vectorial es equivalente a la solución de un modelo de programación cuadrática con restricciones lineales.

#### **4.4.2 Objetivos de la Máquina de Soporte Vectorial**

Para (Tibshirani y Friedman 2012) Los clasificadores de margen máximo constituyen la base teórica sobre la que se construye el concepto de máquina de soporte vectorial, el cual finalmente es la combinación de este tipo de clasificadores de margen máximo con una función Kernel. El objetivo de un clasificador de margen máximo es encontrar una superficie de decisión o hiperplano ( $\mathbf{w}^T \mathbf{x} + b = 0$ ) que sea equidistante a las fronteras de las clases, justo en aquel lugar en el que las clases se encuentran más próximas y maximizar el margen existente entre estas clases. Como se puede observar en la figura

Bajo un enfoque matemático el objetivo de la máquina de soporte vectorial para (Maldonado y Webber 2012) es encontrar, entre todos los hiperplanos que clasifican correctamente los datos, aquel que tenga un mínimo norma  $\|W^2\|$ , lo que equivale maximizar una determinado distancia o Margen existente entre las fronteras que delimitan ciertas clases, en el caso linealmente separable. Cuando se tenga un caso no linealmente separable se procede a construir un hiperplano en un espacio de dimensionalidad muy alta que separe las clases que tenemos mediante una transformación no lineal.

Se puede concluir que el objetivo principal de la Máquina de Soporte Vectorial es Establecer los procedimientos para clasificar objetos (individuos, empresas etc.) al igual que el análisis discriminante de Fisher, pero que se diferencia de esta por pertenecer al campo no paramétrico que incorpora un algoritmo de aprendizaje computacional en su estructura.

#### 4.4.3 Funciones kernel

En el tratamiento de datos es conveniente muchas veces, antes de iniciar un análisis, realizar una representación adecuada de ellos, por ejemplo en escenarios complejos es más conveniente presentar los datos como vectores multivariados: es decir dado un conjunto de datos  $X$  (espacio de entrada), se puede asignar un  $\varphi$ , de tal manera que  $\varphi(x)$  sea la representación vectorial de  $X$  (espacio de características). La asignación o transformación permite analizar instancias de datos complejos a través de métodos de análisis numéricos.

$$x: \mathcal{R}^n \rightarrow \varphi(x): H$$

Las funciones asociadas con estas transformaciones o asignaciones se denominan funciones kernel llamado también por algunos autores como Truco del Kernel, donde la principal ventaja es que se puede aplicar Métodos de análisis lineal en el espacio de características. Sin embargo, debido a que las características son combinaciones no lineales de los datos originales, esto nos permite explorar patrones y relaciones no lineales.

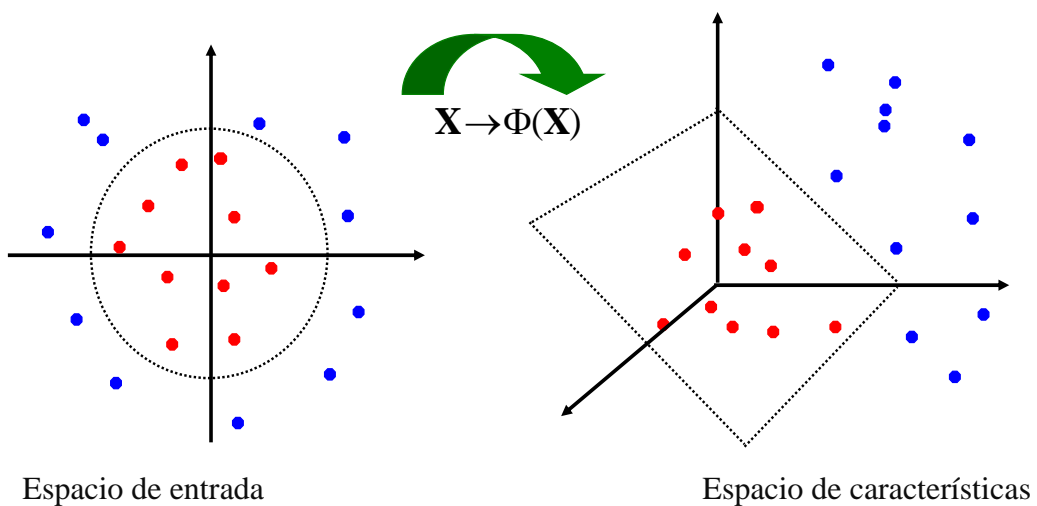
Martínez (2008) indica una definición más formal: Existe un producto escalar en el espacio de características que puede ser expresado como función de los datos en el espacio de entrada llamándose a tal producto como kernel y al espacio de mayor dimensión (espacio de características), con lo cual solo se necesita de los productos escalares para reproducir cualquier algoritmo lineal

El producto escalar que expresado como:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{X}_i^T) \phi(\mathbf{X}_j) \quad (4.7)$$

Un ejemplo de la funcionalidad de funciones kernel se muestra en el siguiente gráfico:

Gráfico N° 1: Noción de una función Kernel



Espacio de entrada

Espacio de características

Fuente: Elaboración Propia

### Tipos de funciones Kernel

Dentro de las principales funciones kernel existentes se tiene:

1. Kernel Lineal:  $K(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i^T \times \mathbf{X}_j$

2. Kernel de base radial (RBF):  $K(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2\rho^2}\right)$   
 $\equiv \exp(-\gamma\|\mathbf{X}_i - \mathbf{X}_j\|^2)$  donde  $\rho, \gamma > 0$  es el parámetro que controla el ancho del kernel.
3. Kernel Tangencia= $Tan(\gamma X_i^T X_j)$ ,  $\gamma > 0$

#### 4.4.4 Clasificador máximo margen

Sea el conjunto de datos de entrenamiento o aprendizaje conformado de n pares:

$$(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_n, y_n) \quad \mathbf{X}_i \in R^n \quad y_i \in \{-1, 1\}$$

Se desea obtener una ecuación (4.6) para un hiperplano para separar dicho grupo de entrenamiento, de manera que aquellos puntos con igual clase queden al mismo lado del hiperplano mientras se maximiza la distancia de dichas clases al hiperplano (Gráfico 3).

$$\mathbf{w}^T \mathbf{x}_i + b \begin{cases} \geq 1 & \text{si } y_i = +1 \\ \leq -1 & \text{si } y_i = -1 \end{cases} \quad \equiv \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \forall i = 1, \dots, n \quad (4.8)$$

Donde  $\mathbf{W}$  es el vector peso y  $b$  es el umbral.

Sea la ecuación de dicho plano o superficie de decisión:  $\mathbf{w}^T \mathbf{x} + b = 0$ . Se toma 2 puntos del plano como se observa en gráfico 3. Sea  $\mathbf{X}_1$  y  $\mathbf{X}_2$  estos puntos que se encuentran en el hiperplano óptimo y satisfacen dicha ecuación del plano:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_1 + b = -1, \quad \mathbf{w}^T \mathbf{x}_2 + b = 1 \quad \mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_2) = 2 \\ \frac{\mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_2)}{\|\mathbf{W}\|} = \frac{2}{\|\mathbf{W}\|} \end{aligned} \quad (4.9)$$

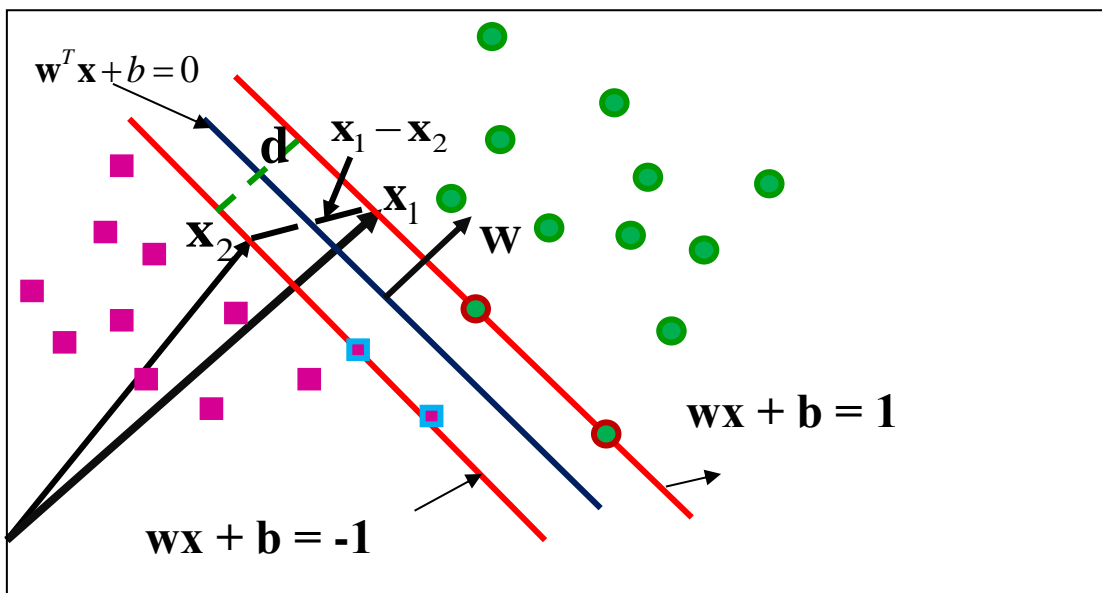


Dado que nuestro interés es que esta Margen sea máximo, ya que de esta forma permitirá distinguir de forma más clara las regiones donde caen los puntos con distintas etiquetas (Abril, 2003) entonces el problema es abordado como un problema de optimización sujeta a las restricciones lineales.

### Caso: Poblaciones Separable Linealmente

Se presenta en casos en los cuales los datos de entrenamiento son linealmente separables, es decir, existe un hiperplano que clasifica perfectamente cada punto como se observa en el gráfico N° 2.

Gráfico N° 2 Poblaciones separable-linealmente



Fuente: Elaboración Propia

La tarea es encontrar el hiperplano que maximice el margen, sujeto a las  $n$  restricciones: Este es un problema clásico de optimización cuadrática clásico con restricciones de desigualdad (Wang, 2005)

$$\begin{aligned}
& \text{Max } \frac{2}{\|\mathbf{W}\|} \equiv \text{Min } \frac{1}{2} \mathbf{W}^T \mathbf{W} \\
& \text{S.a } y_i(\mathbf{W}^T \mathbf{x}_i + b) \geq 1 \quad \forall i, \dots, n
\end{aligned} \tag{4.10}$$

Esta formulación es conocida como problema de optimización primal, Sin embargo es común resolverlo como un problema dual el cual se obtiene mediante el uso de los multiplicadores Lagrange. Para esto se introduce un multiplicador Lagrange  $\alpha_i$  para cada restricción, basado en las condiciones de KKT (Mohammed y Wagner, 2014).

$$\alpha_i \left( y_i(\mathbf{W}^T \mathbf{x}_i + b) - 1 \right) = 0, \quad \alpha_i \geq 0$$

Incorporando para todas las  $n$  restricciones, la nueva función objetiva (4.9) es llamada langrangiano. Donde se tomará derivadas con respecto a  $\mathbf{W}$  y  $b$  para obtener la formulación dual, para ser minimizado respecto a estas mismas variables.

$$\min L_p(\mathbf{w}, b, \alpha_i) = \frac{1}{2} \mathbf{W}^T \mathbf{W} - \sum_{i=1}^n \alpha_i \left( y_i(\mathbf{W}^T \mathbf{x}_i + b) - 1 \right) \tag{4.11}$$

Donde  $\alpha_i$  son los multiplicadores no negativos de Lagrange.

$$\frac{\partial L_p}{\partial \mathbf{W}} = 0 \rightarrow \mathbf{W} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \tag{4.12}$$

$$\frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \tag{4.13}$$

Sustituyendo (4.10) y (4.11) en (4.9) se obtiene la nueva función objetivo Lagrangiano dual (4.12) o formulación dual que no depende de  $\mathbf{W}$  ni  $b$  sino que se especifica sólo en términos de los multiplicadores de Lagrange. Se observa que este lagrangiano dual esta expresado en términos de la data de entrenamiento y depende únicamente del

producto escalar de los patrones de entrada  $(\mathbf{x}_i^T \mathbf{x}_j)$  (Wang, 2005), Este Concepto es de suma importancia para definir luego a la Máquina de soporte vectorial (Webb y Copsey, 2004).

Para encontrar el hiperplano optimo se debe se debe maximizarse respecto a  $\alpha_i$  sujeto a las restricciones.

$$L(\alpha_i) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (4.14)$$

S.a(KKT)  $\alpha_i \geq 0$ , y  $\sum_{i=1}^n \alpha_i y_i = 0$

Esta nueva formulación dual puede ser abordado como un problema programación cuadrática convexa (en base a  $\alpha_i \alpha_j$  términos) donde en muchos casos, puede converger de manera eficiente para el óptimo global mediante la adopción de algunas técnicas de optimización apropiados, tales como el algoritmo de optimización mínima secuencial (SMO) Cristianini et al, (2000) sin embargo existe muchos algoritmos que se pueden aplicar para dar solución a este problema de la programación cuadrática.

De la condición complementaria de KKT ( $\alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1) = 0$ ) podemos desprender que los  $\alpha_i$  que cumplen con ser mayores a cero son llamados los vectores soporte  $\alpha_i \neq 0$ , y serán los únicos que participan en la construcción del hiperplano de clasificación (Maldonado y Weber ,2012) .

Una vez obtenido  $\alpha_i$  para  $1, \dots, n$ , obtenemos los valores de  $\mathbf{W}$  y  $b$  usando (4.6).

$$\mathbf{w} = \sum_{i=1, \alpha_i \neq 0}^n \alpha_i y_i \mathbf{x}_i \quad , \quad b_i = \frac{1}{y_i} - \mathbf{w}^T \mathbf{x}_i = y_i - \mathbf{w}^T \mathbf{x}_i \quad (4.15)$$

Se observa que ya se ha estimado todo los parámetros, finalmente la función de decisión será (Shigeo, 2005):

$$h(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (4.16)$$

Donde  $S$  indica al conjunto de los vectores soporte, los cuales son relevantes para el cálculo del hiperplano óptimo [38]. Por lo tanto un nuevo dato  $\mathbf{X}$  se clasificara dentro sí.

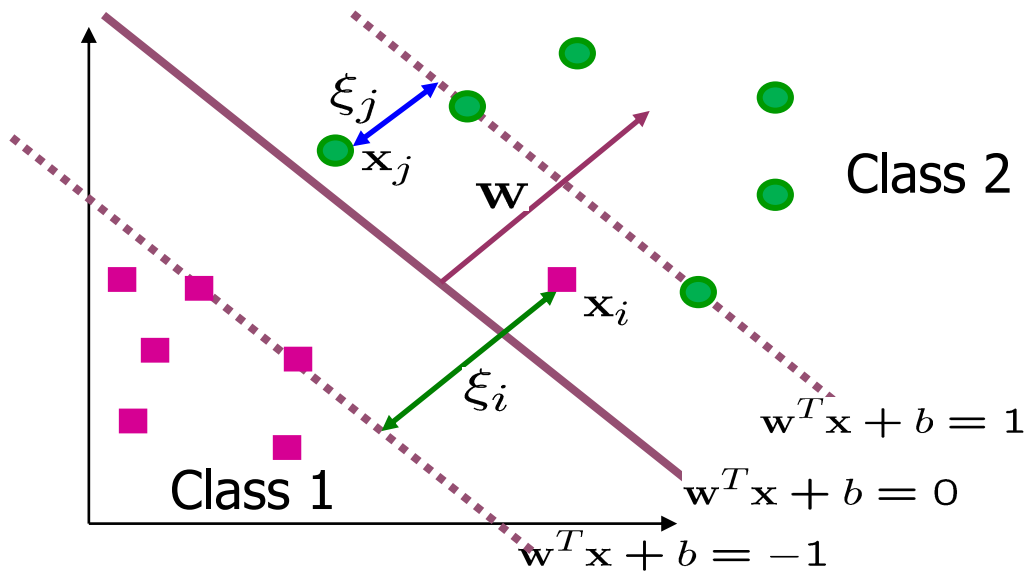
$$h(\mathbf{X}) \begin{cases} > 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases} \quad (4.17)$$

### **Caso: Poblaciones No Separable Linealmente**

En el caso anterior se asumió que los datos de entrenamiento eran linealmente separables. Cuando los datos no son linealmente separables, como ocurre en la mayoría de los casos reales no hay solución factible y el clasificador anterior es insoluble; quedando algunos puntos queden mal asignados a un grupo, estos puntos representarían un error de entrenamiento para el caso anterior. En este caso se introducen variables de holgura,  $\xi$  que cuantifican la distancia de un punto que se encuentra del lado equivocado de la frontera de separación como se observa gráfico N° 3.

En otras palabras un variable de holgura se introduce para dar cuenta de la cantidad de violación de la clasificación por el clasificador (Wu y Kumar, 2009):

Gráfico N° 3 poblaciones no separable linealmente



Fuente: Elaboración Propia

De manera equivalente que en el apartado anterior, la función objetivo se plantea pero esta vez se penaliza para casos mal clasificados donde cada variable de holgura, una por punto, mide la distancia de un punto al margen como  $\xi_i$  o  $\xi_j$  según que corresponde al grupo al que pertenece el punto. De acuerdo con la figura superior.

$$\begin{aligned} \text{Min}_{w,b,\xi_i} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \left( \sum_{i=1}^N \xi_i^k \right) \\ \text{s.a.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall x_i, \quad \xi_i \geq 0 \end{aligned} \quad (4.18)$$

El primer término es minimizado para controlar la capacidad de aprendizaje del mismo modo que el caso lineal; el segundo término permite mantener bajo control el número de clasificaciones erróneas. El parámetro C es elegido por el usuario de manera que un valor grande es equivalente a asignar una alta penalización a los errores. El Parámetro C para Hamel (2009):

“La constante C, llamado el costo, nos permite controlar el equilibrio entre el tamaño del margen y el error (penalizar errores). C es positiva y no puede ser cero; es decir, no

se puede ignorar las variables de holgura ajustando  $C = 0$ . Con un valor grande para  $C$ , la optimización va a tratar de encontrar una solución con el menor número de variables de holgura distintos de cero como sea posible porque los errores son costosos, debido a un gran  $C$ . Un valor grande para  $C$  obliga a la optimización para considerar soluciones con pequeños márgenes.”

En el caso de especificar un valor pequeño para  $C$ , la introducción de variables de holgura no nulos es mucho más tolerante y podemos encontrar soluciones con un margen mayor, haciendo caso omiso de algunos de los puntos más ruidosos cerca de la superficie de la decisión.

Esto da la siguiente relación entre el tamaño y el margen de

$$\begin{aligned} \uparrow C : & \text{Margen pequeño} \\ \downarrow C : & \text{Margen Grande} \end{aligned}$$

De la última expresión (4.16) Asumiendo  $k = 1$ . Se puede calcular el langrangiano para el problema de optimización introduciendo multiplicadores de langrange no negativos  $\alpha_i$  y  $\beta_i$  ( $\beta_i$  se introducen para garantizar positividad de  $\xi_i$ ):

$$\begin{aligned} \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi) &= 0 \text{ Con } \alpha_i \geq 0 \\ \beta_i (\xi_i - 0) &= 0 \text{ Con } \beta_i \geq 0 \end{aligned} \tag{4.19}$$

El langrangiano es dado por:

$$\min L_p(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1) + C \left( \sum_{i=1}^N \xi_i \right) - \left( \sum_{i=1}^N \beta_i \xi_i \right) \tag{9.1}$$

Tomando derivadas parciales respecto a  $\mathbf{w}, b$  y  $\xi_i$

$$\begin{aligned}\frac{\partial L_p}{\partial \mathbf{w}} = 0 &\rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\ \frac{\partial L_p}{\partial b} = 0 &\rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L_p}{\partial \xi} = 0 &\rightarrow C - \alpha_i - \beta_i = 0\end{aligned}\tag{4.20}$$

Sustituyendo estos últimos valores obtenidos en la función objetivo se obtiene la formulación dual que no depende de  $\mathbf{w}$  ni  $b$ . Se debe maximizar respecto a  $\alpha_i$  y sujeto a las restricciones.

$$\begin{aligned}\text{maximize } L(\alpha_i) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.a. } 0 \leq \alpha_i \leq C &\quad \sum_{i=1}^n \alpha_i y_i = 0\end{aligned}\tag{4.21}$$

Una vez resuelto  $\alpha_i$ , se tiene una situación una situación similar que al anterior apartado.

$$\mathbf{w} = \sum_{i=1, \alpha_i \neq 0}^n \alpha_i y_i \mathbf{x}_i, b_i = \frac{1}{y_i} - \mathbf{w}^T \mathbf{x}_i = y_i - \mathbf{w}^T \mathbf{x}_i\tag{4.22}$$

Para cada vector soporte  $\mathbf{x}_i$ , tal que  $0 \leq \alpha_i \leq C$

Una vez que el óptimo hiperplano ha sido determinado, La técnica del SVM puede predecir a clases para nuevos puntos

$$h(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \sum_{i \in S} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b$$

Donde  $S (\alpha_i > 0)$  indica al conjunto de los vectores soporte. Por lo tanto un nuevo dato  $\mathbf{X}$  se clasificara dentro si.

$$h(\mathbf{X}) \begin{cases} > 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

(4.23)

#### 4.4.5 Construcción de la Máquina de Soporte Vectorial

Se ha presentado hasta ahora a los clasificadores en sus diversas formas basados para reglas de clasificación lineales, existe un hiperplano separador, el cual es inapropiado para muchos Problemas del mundo real ya que estos tiene una estructura intrínsecamente no lineal (Joachims, 2001). Una notable propiedad de la Máquina de Soporte vectorial es que fácilmente se puede transformar dentro de un marco aprendizaje no lineal.

La idea fundamental es transformar el espacio de entrada (gráfico N° 2) en el que el conjunto de datos no es linealmente separable en un espacio de Alta dimensión llamada espacio de características, donde los datos son linealmente separables, una vez aquí se construye un hiperplano separador de máximo margen. Cabe destacar que si elegimos estas transformaciones cuidadosamente, todos los cálculos relacionados con el espacio de características se pueden realizar en el espacio de entrada. Es decir, a pesar de que estamos transformando nuestro espacio de entrada para que los datos se convierten linealmente separables, no tenemos que pagar el coste computacional de estas transformaciones (Hamel 2009), Las funciones asociadas con estas transformaciones se denominan funciones de núcleo o kernel, y el proceso de utilizar estas funciones para moverse de forma lineal a una máquina de vectores de soporte no lineal se llama el truco del Kernel



La selección del mejor kernel es todavía un tema de investigación como refiere (Ali y Smith, 2006). El procedimiento más común es el de seleccionar los parámetros de kernel es calibrando estos parámetros en conjunto con el proceso de selección de la mejor regla (parámetro C que controla la generalización de la regla de clasificación) mediante una búsqueda en rejilla (Hsu et al 2003).

Retomando el camino para definir La Máquina de Soporte Vectorial: (Morales, 2009) Afirma al Truco del Kernel como una sustitución de los productos escalares que nos da una medida de distancia en términos de los kernel vistos a las ecuaciones anteriormente vistas en los apartados.

Esto es:

$$(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_N, y_N) \rightarrow (\phi(\mathbf{X}_1), y_1), (\phi(\mathbf{X}_2), y_2), \dots, (\phi(\mathbf{X}_N), y_N) \quad y_i \in \{-1, 1\}$$

Aplicando  $\phi$  a cada punto la función objetivo:

$$\begin{aligned} \text{Min} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \left( \sum_{i=1}^N \xi_i^k \right) \\ \text{S.a} \quad & y_i (\mathbf{w}^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \forall x_i, \quad \xi_i \geq 0 \end{aligned} \tag{4.24}$$

Como se puede observar en la formulación dual de los problemas de clasificación mediante margen máximo (tanto SVM lineal como SVM no lineal) los elementos que componen la muestra de aprendizaje aparecen únicamente expresados en función de sus productos escalares  $\mathbf{x}_i^T \mathbf{x}_j$  y esto se puede sustituir por una función kernel  $\phi(\mathbf{x}_i^T) \phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$  donde esto vendría a la razón de ser de una Support Vector Machine.

Análogamente de la función objetivo se desprende la forma Primal el cual sigue el mismo procedimiento para encontrar la siguiente forma dual para ser maximizado.

$$\begin{aligned} \max L(\alpha_i) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.a.} \sum_{i=1}^n \alpha_i y_i &= 0, \quad 0 \leq \alpha_i \leq C \end{aligned} \quad (4.25)$$

A causa de que  $K(\mathbf{x}_i, \mathbf{x}_j)$  es un kernel semidefinido, el problema de optimización es un problema de optimización cuadrática. Resolviendo  $\mathbf{W}$  en el espacio de características de la siguiente manera:

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^n \alpha_i y_i \phi(\mathbf{x}_i), \quad \alpha_i \geq 0 & b &= \frac{1}{n_{sv}} \sum_{\alpha_i > 0} y_i - \sum_{\alpha_i > 0} \sum_{\alpha_j > 0} \alpha_i y_i \phi(\mathbf{x}_i^T) \phi(\mathbf{x}_j) \end{aligned} \quad (4.26)$$

Donde  $n_{sv}$  es el número de vectores soporte con  $\alpha_i > 0$ . Entonces podemos predecir la clase para un nuevo punto  $\mathbf{x}_i$  con:

$$h(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (4.27)$$

Donde  $S (\alpha_i > 0)$  indica al conjunto de los vectores soporte.

Un dato desconocido será clasificado usando la función:

$$h(\mathbf{X}) \begin{cases} > 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$

Como se ha visto en cada uno de los casos, La estimación de los parámetros de una SVM es equivalente a la solución de un modelo de programación cuadrática con restricciones lineales; ello implica, que la solución óptima es global y única, lo cual representa una clara ventaja para esta técnica (Velásquez et al, 2010).

## **Algoritmo Máquina Soporte Vectorial**

Existen varios algoritmos para resolver el problema de la programación cuadrática visto en los distintos apartados anteriores en la formulación dual: Entre ellos Chunking (Vapnik, 1982), Sequential Minimum Optimization (Platt,1998), Osuna-SVMLight (Osuna, 1997).

Chunking está limitada por el número máximo de vectores de soporte que se pueden manejar, esto consiste en tomar un conjunto arbitrario inicial de vectores a partir de los cuales se entrena la SVM. El algoritmo de (Osuna, 1997) descompone el problema grande de QP en una serie de pequeños problemas de QP, enfoque que (Platt, 1998) utilizaría para Presentar al Algoritmo SMO. De estos algoritmos Presentados SMO es el de mayor impacto ya que logra mejorar la rapidez del entrenamiento de la SVM, que implica resolver sub problemas del problema inicial para que el coste computacional sea menos costoso. Es más fácilmente programable, no requiere de cálculos complejos para resolver el problema de programación cuadrática que SVM plantea donde todo esto representa sus principales ventajas (Platt, 1999) Así este autor indica: "Hay dos componentes en un SMO: un método analítico para la solución de los dos multiplicadores de Lagrange, y una heurística para elegir qué multiplicadores para optimizar".

En esta investigación se usará La Máquina de Soporte Vectorial bajo el algoritmo Sequential Minimum Optimization (Platt, 1998)

### **4.4.6 Pasos de la metodología de la Máquina de Soporte Vectorial**

#### **Paso 1: Objetivos**

El objetivo Principal de la Máquina de soporte vectorial es encontrar un encontrar una superficie de decisión o hiperplano con máximo margen que delimite las fronteras que delimitan ciertas clases, en el caso linealmente separable. Cuando se tenga un caso no linealmente separable se procede a construir un hiperplano en un espacio de

dimensionalidad muy alta que separe las clases que tenemos mediante una transformación no lineal llamado Truco del Kernel. Este objetivo con lleva a establecer una regla o técnica con los parámetros adecuados propios del Kernel y de la técnica misma para finalmente realizar las clasificaciones posteriores.

Finalmente se puede concluir que el objetivo principal de la Máquina de Soporte Vectorial es establecer los procedimientos para clasificar objetos (individuos, empresas etc.) al igual que el análisis discriminante de Fisher, pero que se diferencia de esta por pertenecer al campo no paramétrico que incorpora un algoritmo de aprendizaje computacional en su estructura, donde los supuestos que requiere para su ejecución son Bastante tolerables.

## **Paso 2: Diseño de la investigación**

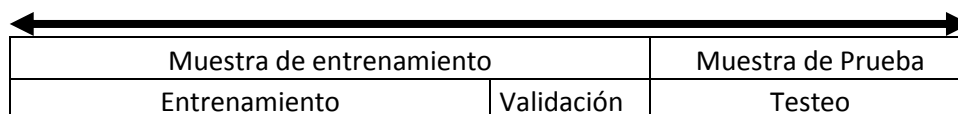
En el segundo paso el investigador debe tomar decisiones sobre que variable deberá ser considerada como dependiente y que variables como independientes, sobre el tamaño de la muestra que va a considerar en el estudio y en la estrategia para dividir la muestra con fines de estimación de la función discriminante y otra con fines de validación.

En las investigaciones del campo del aprendizaje automático es común que usar tres conjuntos de datos independientes como se puede apreciar en el cuadro N° 1.

- a) Muestra de entrenamiento (Training) que consta de datos con los que se entrenan a fin de encontrar la mejor regla de clasificación.
- b) Muestra de validación (Validation) que selecciona el mejor de las reglas de clasificación que se ha entrenados y finalmente:
- c) Muestra de prueba (Test): entrega el error real cometido con la regla de clasificación seleccionado.

Es importante que el conjunto de entrenamiento y validación sean del mismo conjunto de datos inicial es decir de la misma población, pero el conjunto de prueba puede o no salir del conjunto de datos inicial.

Cuadro N° 1: Partición data total



Muestra de entrenamiento		Muestra de Prueba
Entrenamiento	Validación	Testeo

La metodología para estas 3 muestras independientes es que la máquina de soporte vectorial se ajustan a los datos de entrenamiento y el error de clasificación para dicha regla de clasificación es obtenido mediante el uso de los datos de validación. Este error de clasificación en los datos de validación se puede utilizar (es decir, el algoritmo los utiliza) para decidir cuándo dar por terminado el proceso de selección o para decidir cuáles son los efectos a incluir a medida que avanza el proceso. Finalmente, una vez que termina el proceso y se tiene seleccionado la mejor regla de clasificación, se pueden utilizar los datos de prueba para evaluar la manera en que la regla de clasificación seleccionado se generaliza para los datos que no jugaron ningún papel en la selección del mismo.

Hastie et al (2011) señalan que es difícil dar una regla general sobre cuántas observaciones se deben asignar a cada conjunto, aunque indican que una división típica puede ser de 50% para el entrenamiento y 25% para la validación y prueba, respectivamente. Para dividir la muestra Siddiqi (2012) propone una muestra 70% para el desarrollo de la regla de clasificación y el 30% restante para la prueba

### Paso 3: Supuestos

La Máquina de Soporte requiere que los datos de entrada estén representados como un vector de números reales. Por tanto, de existir atributos no numéricos, estos deben transformarse en datos cuantitativos. También es recomendable escalar los datos antes de aplicar La Máquina de Soporte Vectorial. La principal ventaja del escalamiento es evitar que los atributos expresados en datos numéricos grandes dominen aquellos expresados en datos numéricos pequeños Hsu et al (2016).

Es importante realizar el diagnostico de multicolinealidad de la misma forma que se hace en el Análisis Discriminante de Fisher. La Máquina de Soporte Vectorial al tratarse

de una técnica computacional que pertenece al aprendizaje automático es bastante tolerante con los datos de entrada, especialmente con la versión de margen suave.

#### Paso 4: Estimación de la Máquina de Soporte Vectorial

Búsqueda y selección de los mejores parámetros:

En el aprendizaje automático es muy importante seleccionar bien los parámetros de esta manera se asegura una buena generalización de los datos con la estimación de la regla de clasificación para la Máquina de soporte vectorial.

De la relación (4.23)

$$\max L(\alpha_i) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad \text{s.a.} \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C$$

De la expresión última, a causa de que  $K(\mathbf{x}_i, \mathbf{x}_j)$  es un Kernel, en el caso de una función Kernel de base radial, existen 2 parámetros: *costo*( $C$ ), y *gamma* ( $\gamma$ ), estos parámetros son determinados de manera experimental al realizar el proceso de modelamiento con la data aprendizaje. Como se ha explicado  $C$  es un parámetro de costo del clasificador de Máquina de soporte vectorial, el cual permite mayor flexibilidad en la separación de clases y por un ende una mayor generalización.  $\gamma$  Es un parámetro propio de la función Kernel llamado Gamma, y su rol es de escalar la información de las observaciones sobre el espacio de la transformación Kernel tiene un valor por defecto que se calcula como el cociente entre  $\frac{1}{\text{Dimensión datos}}$ .

Todos los Kernel necesitan el parámetro *gamma* ( $\gamma$ ), excepto el kernel lineal.

Para encontrar o calibrar estos dos utilizando función Kernel se empieza usando un conjunto de combinaciones de estos, debido a que sus valores no son conocidos previamente, para su asignación y mejor combinación se realiza una Búsqueda

Exhaustiva en Rejilla (Grid-Research) con secuencia de crecimiento exponencial para  $C$ ,  $\gamma$ .

(Moro, 2007) usa un procedimiento experimental para encontrar los parámetros:

- 1) Primero usar una búsqueda en rejilla con valores posibles con pares de  $\text{costo}(C)$ , y  $\text{gamma}(\gamma)$ , según el tipo de Kernel, donde se busca lo mejores parámetros que clasifiquen con precisión.
- 2) Usar una búsqueda de rejilla más fina sobre dicha zona
- 3) Se entrena de nuevo todo el conjunto de entrenamiento con el mejor par (Costo, gamma) para obtener el clasificador final.

El proceso anterior es sintetizado por (Hsu et al 2000) donde emplea la búsqueda por Rejilla dando valores del Costo  $C$  y de la Gamma,  $\gamma$ , usando validación cruzada. Se utiliza secuencias crecientes de valores exponenciales y se realizan dos búsquedas: en la primera se intenta identificar la mejor región –la región determinada por los  $C$  y  $\gamma$  que producen la mejor probabilidad de acierto –mientras que en la segunda se busca sobre esa región el par  $(C, \gamma)$  que entrega el mejor resultado.

Gareth et al. (2013) y Meyer et al. (2017) utilizan como valores de inicio para la red de búsqueda para  $\text{Gamma} = 2^{(-1:1)}$  y para  $\text{Costo} = 2^{(2:4)}$  los cuales ingresan a la red de búsqueda que mediante validación cruzada arrojan un performance para dicha regla de clasificación con tales parámetros.

Existen algunos procedimientos alternativos para encontrar los mejores parámetros, alguno de ellos de manera automática que contiene en su estructura valores por defecto, Meyer (2004) recomienda no utilizar estos valores dados por defecto, pues son poco probables que ofrezcan reglas de clasificación con buen rendimiento.

## **Paso 5: Interpretación de los resultados**

El número de vectores de soporte, se definen como aquellos vectores que se encuentra en el Margen del Hiperplano que separa a cierta clase de la otra. Para un conjunto de variables de un problema específico, una cantidad elevada de número de vectores de soporte puede significar que no ha conseguido distinguir los rasgos comunes de cada clase porque las variables empleadas no son las adecuadas y, por lo tanto, se aprende “de memoria” los datos de entrada lo que puede conllevar a un sobreajuste. De igual forma un número elevado de vectores soporte significa que el conjunto de variables sirven para caracterizar los datos, pero la forma de su distribución dificultan la delimitación de las clases, requiriéndose más puntos para poder encontrar el hiperplano.

Observar los parámetros óptimos resultantes, luego de entrenar cada diferente regla de clasificación es de suma importancia, en el caso de la generación de la regla de clasificación a través de la Máquina de Soporte vectorial mediante un Kernel Lineal el parámetro a utilizar es el parámetro  $C$ , llamado Coste. En el caso del Kernel Radial o Gaussiano, además del parámetro Costo, el parámetro sigma controla la anchura de la campana de Gauss. Con valores grandes de sigma, los puntos de nuestras características conocidas podrán tomar más valores en su dominio, teniendo así cambios más suaves en sus valores. Por el contrario, con valores pequeños tendremos un dominio de valores más reducido, con más varianza.

La elección y afinación del número de vectores de soporte y los óptimos parámetros en cada modelamiento de la Máquina de Soporte vectorial en función a cada Kernel se puede medir en base a los indicadores de la validación Cruzada.

## **Paso 6: Validación de los resultados**

La validación de los resultados es la última etapa de su aplicación y permite asegurar que los resultados tengan validez externa e interna. Los procedimientos de mayor uso para verificar la validez de los resultados es la validación cruzada, donde se dividen grupos aleatoriamente para luego ampliar dicha muestra para verificar su capacidad clasificatoria.



#### 4.5. Evaluación de la clasificación

Para (Díaz et al, 2014). “No existe una regla de clasificación mejor que otro de manera general; para cada problema nuevo es necesario determinar con cuál se pueden obtener mejores resultados, y es por esto que han surgido varias medidas para evaluar la clasificación y comparar las técnicas empleadas para un problema determinado”.

Las medidas más conocidas para evaluar la clasificación están basadas en la matriz de confusión que se obtiene cuando se prueba el clasificador en un conjunto de datos que no intervienen en el entrenamiento, además los indicadores de la Curva ROC permite evaluar el poder clasificatorio de las técnicas empleadas.

#### Matriz de Confusión

Una matriz de confusión contiene información acerca de las clasificaciones actuales y los previstos o pronosticados por un sistema de clasificación o modelo empleado. El rendimiento de estos sistemas se evalúa habitualmente mediante los datos de una matriz que a través de los indicadores que ofrece la Exactitud, Sensibilidad, Especificidad se pueden tomar decisiones en cuanto al poder clasificatorio de los diversas tecnicas empleadas.

Cuadro N°2: Elementos de la Matriz de Confusión

<b>Matriz de Confusión</b>		<b>Clase Predicha</b>	
		<i>Ingresar</i>	<i>No Ingresar</i>
<b>Clase Verdadera</b>	<i>Ingresar</i>	<i>VP(verdaderos positivos)</i>	<i>FP(falsos positivos)</i>
	<i>No Ingresar</i>	<i>FN(falsos negativos)</i>	<i>VN(Verdaderos negativos)</i>
<b>Total columna</b>		<b><i>P</i></b>	<b><i>N</i></b>

Fuente: Elaboración Propia

De la matriz de confusión para un clasificador de dos clases se pueden obtener los siguientes indicadores: Exactitud, Sensibilidad, Especificidad

Exactitud:

$$exactitud = \frac{VP + VN}{P + N}$$

Es la proporción de resultados válidos entre la totalidad de las pruebas efectuadas con el método de clasificación. El valor que puede asumir la Exactitud varía del 0 al 1 (100%), es decir, cuanto más alto es el valor, hay una mejor capacidad en la clasificación correcta por el método de clasificación.

El complemento de la Exactitud también es llamado ratio de mala clasificación.

Sensibilidad:

$$sensibilidad = \frac{VP}{P}$$

Se define la sensibilidad de una prueba como la proporción de los postulantes que son clasificados como ingresantes en un Concurso de Admisión por ejemplo y que se identifican correctamente con el método de clasificación.

El valor que puede asumir la sensibilidad varía del 0 al 1 (100%), es decir, cuanto más alto es el valor, hay una mejor capacidad en la detección de ingresantes por medio del método de clasificación.

Especificidad:

$$especificidad = \frac{VN}{N}$$

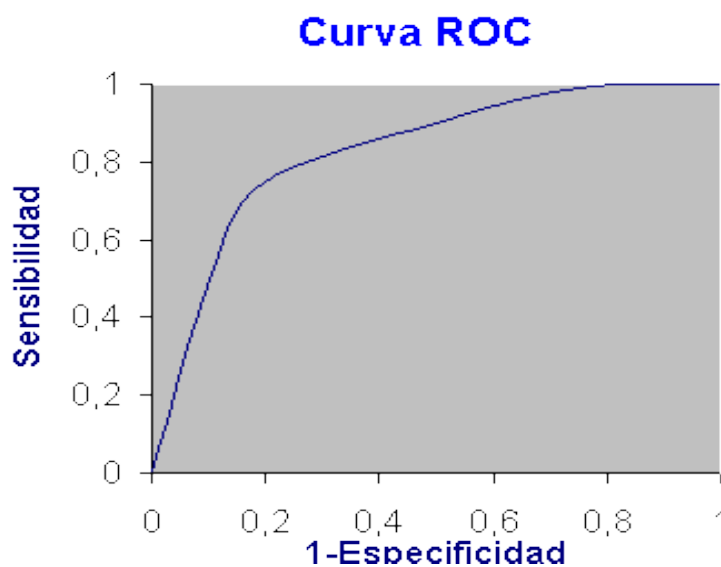
Análogamente al anterior, la especificidad de una prueba en estudio se refiere a la proporción de los individuos Clasificados como No ingresantes en el examen de admisión que se identifican correctamente por de método de clasificación. Al igual que la sensibilidad, el valor de la especificidad varía del 0 al 1 (100%), lo que significa que cuanto mayor sea el valor, mayor capacidad de clasificación como No Ingresantes por el método clasificador.

## Curva ROC

Otra forma de evaluar el rendimiento de un clasificador es por la curva ROC (Receiver Operating Characteristic). Es un método útil para evaluar la realización de esquema de clasificación en los que exista una variable binaria, es decir con 2 clases. La curva ROC se construye a partir de los valores que se obtienen para la sensibilidad y especificidad usando distintos puntos de corte. En el Eje X se sitúa la probabilidad de un falso positivo (complemento de la especificidad), en el Eje Y, la probabilidad de declarar a un verdadero positivo (Sensibilidad). Esto se hace para cada punto de corte que se acoja de 0 a 1.

Es importante observar el área bajo la curva ROC pues cuanto mayor sea esa área (más próxima sea al máximo 1), mayor capacidad predictiva o de clasificación tendrá la función. De la misma forma, el área bajo la curva ROC es una vía para comparar diferentes funciones predictivas. Si una de ellas está por encima de la otra a lo largo del intervalo (0,1), no quedará duda de que el primera regla de clasificación es más eficiente que el segundo. Si ambas curvas se interceptan, entonces se debe definir un indicador que permita decidir. Este indicador global de la eficiencia de la prueba (el área bajo la Curva ROC) que se mueve necesariamente entre 0 y 1

Gráfico N°4: Curva ROC



Fuente: Elaboración propia

## **V. MATERIALES Y MÉTODOS**

### **5.1 Metodología de la investigación**

En la investigación se considera a los postulantes a la UNALM en el Concurso de Admisión 2015-II. Las técnicas Análisis Discriminante Lineal de Fisher y Máquina de Soporte Vectorial se utilizan en dos oportunidades (Aplicación 1 y Aplicación 2) debido a que el objetivo del estudio de comparar su capacidad clasificatoria en los grupos de interés: los que no recibieron preparación en el CEP-UNALM y los que si recibieron preparación en el CEP-UNALM.

En ambos grupos de interés el objetivo es clasificarlos a través de las técnicas aplicadas según los resultados obtenidos en la prueba de admisión de la UNALM (si ingresa y no ingresa) y en función a su rendimiento en las seis materias que se evalúan en la prueba de admisión. Los resultados obtenidos mediante esta estrategia de análisis, permitirá identificar si los postulantes que recibieron preparación en el CEP-UNALM tienen un rendimiento diferenciado en la prueba de admisión de la UNALM de los que no se prepararon en ella y comprobar si efectivamente el CEP-UNALM brinda una formación diferenciada respecto a otras centros de preparación.

#### **5.1.1 Tipo de investigación**

El diseño de la investigación es causal debido a la presencia en cada una de las aplicaciones de una variable dependiente (Y) de naturaleza categórica y un conjunto de variables independientes ( $X_1, X_2, X_3, X_4, X_5, X_6$ ) de naturaleza métrica.

#### **5.1.2 Formulación de hipótesis**

##### **Hipótesis General**

La técnica de Máquina de Soporte Vectorial tiene mejores indicadores en la matriz de confusión que la técnica Análisis Discriminante Lineal de Fisher, cuando se clasifica el

resultado de los postulantes en la prueba de admisión de la UNALM del Concurso de Admisión 2015 –II.

### **Hipótesis Específica**

Con la técnica de Máquina de Soporte Vectorial se tiene un mayor valor de área bajo la curva ROC que con el Análisis Discriminante Lineal de Fisher, cuando se clasifica el resultado de los postulantes en la prueba de admisión de la UNALM del Concurso de Admisión 2015 –II.

Los 6 cursos que se evalúan en el examen del concurso de admisión en la UNALM tienen capacidad discriminante en el perfil del resultado de los postulantes que ingresan o no a la universidad y cuya preparación se realizó o no en el CEP-UNALM

### **5.1.3 Identificación de variables**

En cada aplicación se define la variable dependiente (Y) y las variables dependientes ( $X_1, X_2, X_3, X_4, X_5$  y  $X_6$ ) de la siguiente forma:

#### **Aplicación 1:**

$Y_1$  = resultado en la prueba admisión de la UNALM de los postulantes que no se prepararon en el CEP-UNALM y que ingresaron o no a la universidad.

$X_1$  : Puntaje de Razonamiento Matemático

$X_2$  : Puntaje de Razonamiento Verbal

$X_3$  : Puntaje de Matemáticas

$X_4$  : Puntaje de Física

$X_5$  : Puntaje de Química

$X_6$  : Puntaje de Biología

$$NO\ CEP-UNALM \begin{cases} Ingresó \\ No\ ingresó \end{cases} \text{ vs } X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

### Aplicación 2:

$Y_2$  = resultado en la prueba admisión de la UNALM de los postulantes que si se prepararon en el CEP-UNALM y que ingresaron o no a la universidad.

$X_1$  : Puntaje de Razonamiento Matemático

$X_2$  : Puntaje de Razonamiento Verbal

$X_3$  : Puntaje de Matemáticas

$X_4$  : Puntaje de Física

$X_5$  : Puntaje de Química

$X_6$  : Puntaje de Biología

$$SI\ CEP-UNALM \begin{cases} Ingresó \\ No\ ingresó \end{cases} \text{ vs } X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

#### 5.1.3 Definición operacional de las variables

En las dos aplicaciones la variable dependiente ( $Y$ ) es de naturaleza categórica y las variables independientes  $X_1, X_2, X_3, X_4, X_5, X_6$  son de naturaleza cuantitativa medidas en una escala vigesimal.

#### 5.1.4 Diseño de la investigación

El diseño de la investigación es no experimental de corte transversal. Se considera no experimental porque no hay manipulación de las variables independientes y es de carácter transversal porque se aplica en un momento del tiempo (ciclo 2015-II).

### **5.1.5 Población y muestra**

Población de postulantes al Concurso de Admisión 2015-II de la UNALM.

### **5.1.6 Instrumento de colecta de datos**

El instrumento utilizado en la investigación fue la Prueba de Admisión. Es elaborado por el Comité Permanente de Admisión, su tiempo de duración es de 3 horas y consta de 100 preguntas con cinco alternativas cada una. Las materias evaluadas y el número de preguntas por materia son: Razonamiento Matemático (20), Razonamiento Verbal (20), Matemática (Aritmética (4), Algebra (6), Geometría (4), Trigonometría (4)), Física (14), Química (14) y Biología (14). Cada pregunta bien contestada tiene un valor de 1.00 punto, sin contestar 0.00 y mal contestada – 0.25.

## **5.2 Metodología aplicada**

Antes de aplicar cada técnica, la exploración de las variables o la revisión descriptiva de las variables es un paso crucial, pues es la fase de comprensión de los datos. Se puede denominar este paso como el soporte analítico de las reglas de clasificación, pues brindan una explicación del comportamiento de las variables previo a los resultados de las reglas de decisión.

Ambas técnicas de modelamiento emplean el mismo particionamiento de la data con la finalidad de poder generalizar el patrón que presentan los datos y para que la evaluación de los pronósticos con ambas técnicas se realice en el mismo contexto. Finalmente, para poder evaluar los resultados en cuanto a la capacidad de clasificación de las reglas de clasificación o modelos se separa la data de test o prueba, sobre los cuales serán medidos los indicadores de buena clasificación.

En el ámbito de aplicación de la investigación, aquella que tenga un mayor poder de clasificación en el contexto de ingresante o no ingresantes de acuerdo a la procedencia mostrara mejores resultados

## **VI. RESULTADOS Y DISCUSION**

La comparación de la capacidad clasificatoria de las técnicas del Análisis Discriminante Lineal de Fisher y Máquina de Soporte Vectorial se aplica al rendimiento de los postulantes en la prueba de admisión 2015-II en la Universidad Nacional Agraria La Molina (UNALM). Ambas técnicas estadísticas se aplicaron dos veces. La Aplicación 1 evalúa el rendimiento en la prueba admisión de la UNALM de los postulantes que no se prepararon en el CEP-UNALM y que ingresaron o no a la universidad. La aplicación 2 evalúa el rendimiento en la prueba admisión de la UNALM de los postulantes que si se prepararon en el CEP-UNALM y que ingresaron o no a la universidad.

### **6.1 Resultados aplicación 1**

#### **6.1.1 Exploración de variables:**

El análisis descriptivo y gráfico de las variables independientes (puntajes de los cursos) en cada categoría de la variable dependiente (si ingresó, no ingresó) son los siguientes:

En el Cuadro N° 3 se muestra el puntaje promedio y desviación estándar de cada curso (razonamiento matemático, razonamiento verbal, matemáticas, física, química y biología) en cada condición de ingreso (si ingresó, no ingresó).

En el caso de los ingresantes el puntaje más alto se obtiene en el curso de Química (16.0386) y el más bajo en Razonamiento Verbal (8.6884). En el caso de los no ingresantes el puntaje más alto se registra en Química (8.6984) y el más bajo en Física (4.3322).

En el caso de los ingresantes la variabilidad más alta se presenta en el curso de Biología (3.94392) y el más bajo en Química (2.41555). En el caso de los no ingresantes la variabilidad más alta se presenta en el curso de Química (5.35182) y el más bajo en Razonamiento Verbal (2.94331).



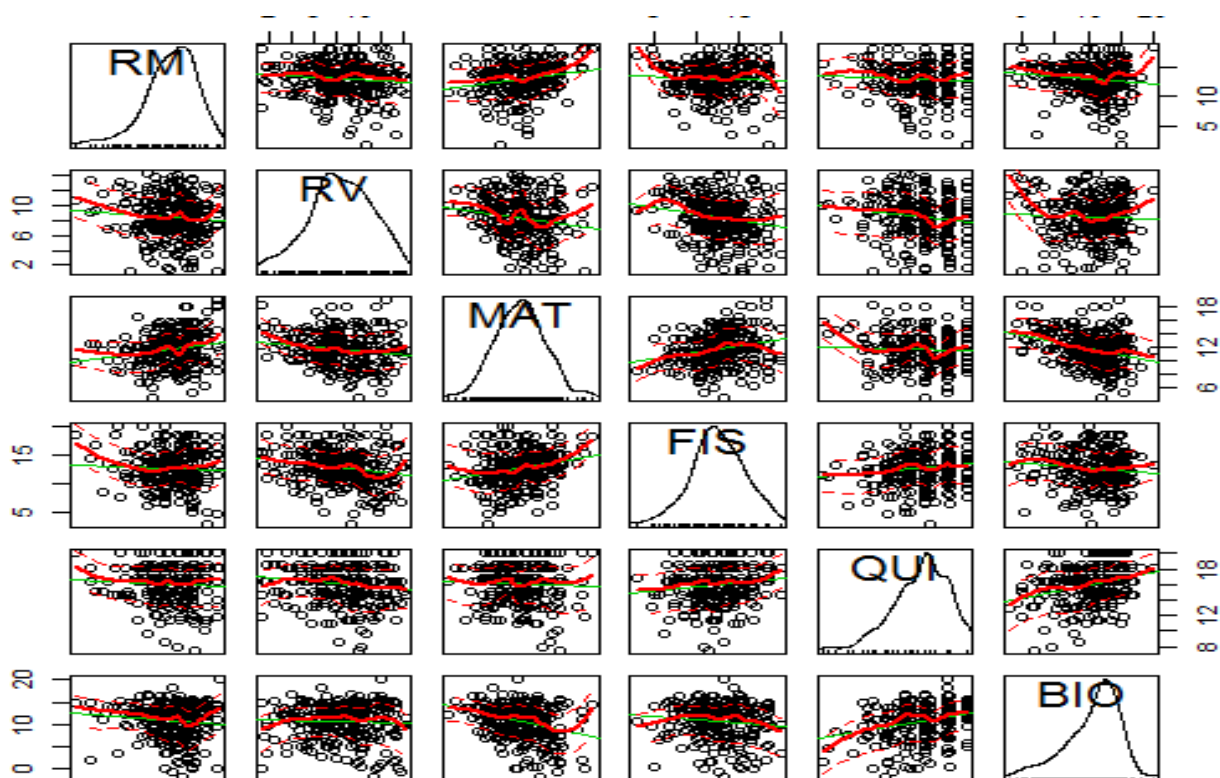
Cuadro N° 3: Resultados descriptivos ingresante y no ingresante

CONDICIÓN	CURSO	MEAN	DESV.
SI INGRESÓ	<i>RM</i>	13.0398	2.89102
	<i>RV</i>	8.3769	2.80803
	<i>MAT</i>	11.6881	2.46079
	<i>FIS</i>	12.6852	3.11774
	<i>QUI</i>	16.0386	2.41555
	<i>BIO</i>	10.7597	3.94392
NO INGRESÓ	<i>RM</i>	7.5773	4.02570
	<i>RV</i>	6.2847	2.94331
	<i>MAT</i>	4.9750	3.52922
	<i>FIS</i>	4.3322	4.32168
	<i>QUI</i>	8.6984	5.35182
	<i>BIO</i>	4.4159	4.26650

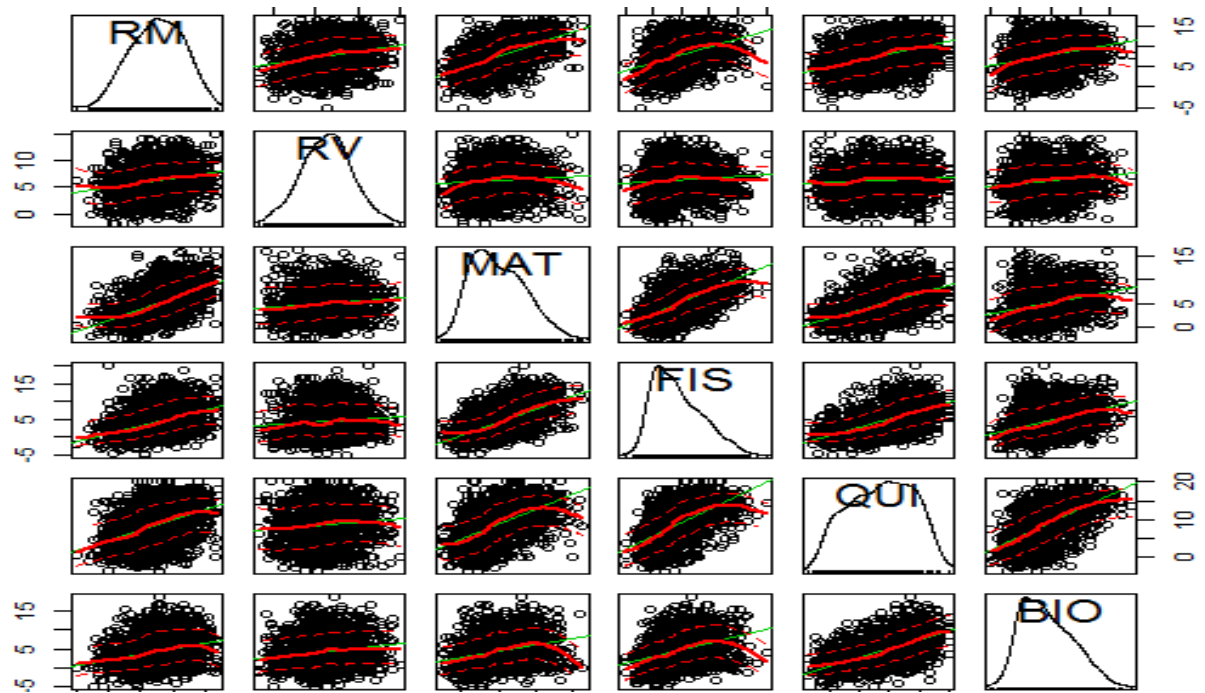
Fuente: Elaboración propia

Gráfico N° 5: Gráfico Descriptivo ingresante y no ingresantes

Si Ingresantes

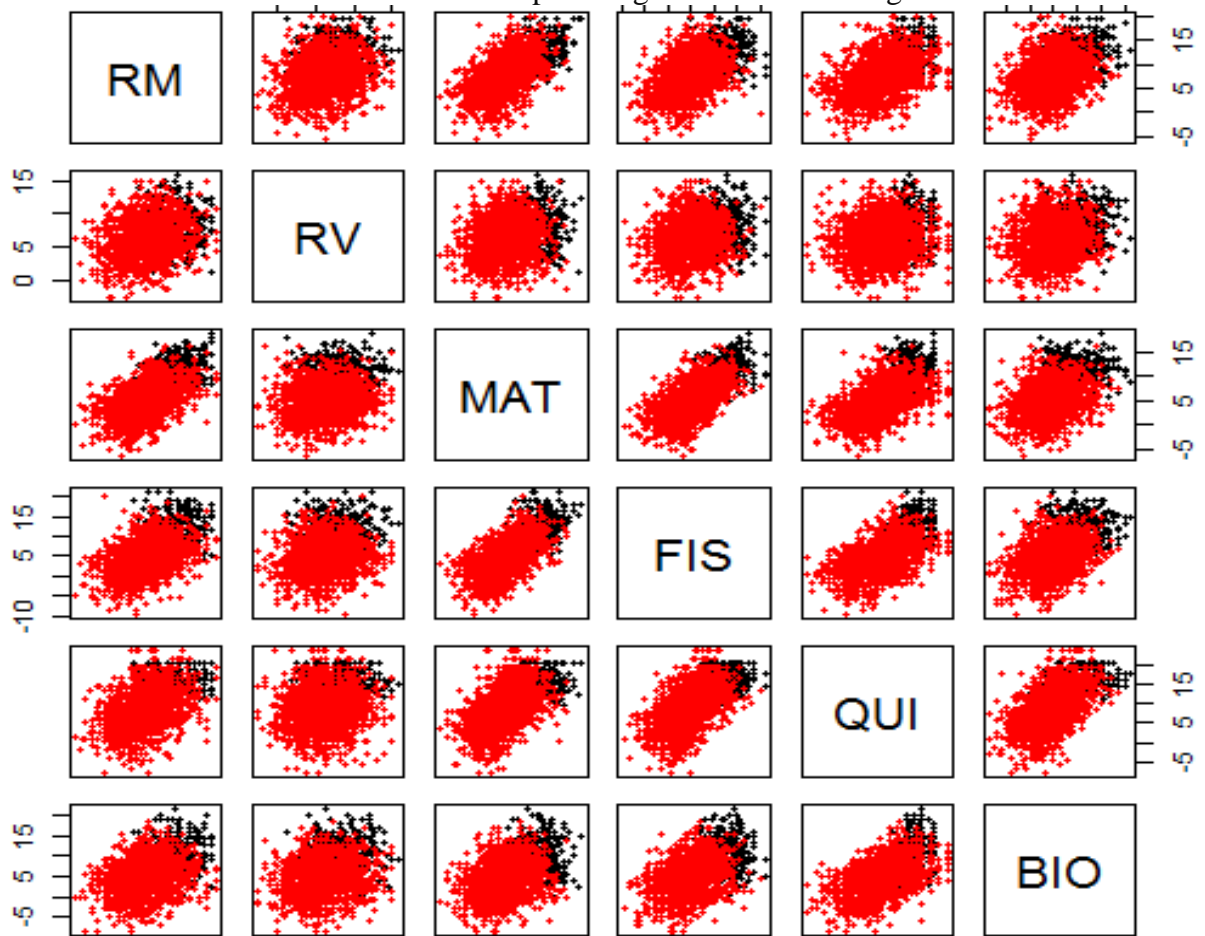


No Ingresantes



Fuente: Elaboración Propia

Gráfico N° 6: Gráfico descriptivo según condición de ingreso



Fuente: Elaboración Propia

En el Grafico N° 5 se observa que no hay simetría en la distribución en las notas de los cursos, por lo que es previsible que no se cumpla el supuesto de normalidad univariada, el cual será verificado posteriormente. Además, se observa una tendencia lineal en la distribución de los cursos lo que permite sospechar la presencia de multicolinealidad, lo que también será verificado posteriormente.

En el Gráfico N° 6 se aprecia que los ingresantes (color negro) y no ingresantes (color rojo) están visiblemente diferenciados en la gráfica de dispersión, por lo que se supone que estarán divididos por un hiperplano lineal, lo cual se comprobará más adelante.

### **6.1.2 Aplicación de las técnicas de clasificación**

#### **Primer Paso: Objetivos**

Uno de los principales objetivos del CEP-UNALM, como institución es que sus alumnos ingresen a la universidad con un rendimiento diferenciado en la prueba de admisión respecto a los postulantes que no se prepararon en ella, demostrando a la sociedad la calidad de los servicios académicos que brinda.

Según la anterior premisa, la técnica Análisis Discriminante Lineal de Fisher tiene como objetivo determinar el perfil en el resultado de los postulantes en las 6 materias de evaluación, es decir, los que ingresaron o no al Concurso de Admisión 2015-II en función a la no preparación en la CEP-UNALM, lo que posteriormente servirá como punto de referencia para su comparación respecto al perfil de los postulantes que si provienen de la CEP-UNALM en la aplicación 2, de la misma forma determinar cuáles de dichas materias evaluadas cuantifica mejor las diferencias en dicho perfil.

En una segunda instancia el objetivo para la técnica de Máquina de Soporte Vectorial y el Análisis Discriminante de Fisher se delimita a encontrar la función de clasificación para los resultados de los postulantes al dicho Concurso de Admisión en tales materias.

## **Segundo Paso: Diseño de la Investigación**

Las variables fueron definidas anteriormente. El tamaño de muestra considerado para la aplicación de ambas técnicas para el grupo de postulantes que ingresaron o no en el Concurso de Admisión y cuya preparación no fue en la CEP-UNALM fue de 1800 postulantes. La estrategia de partición para evaluar el poder de clasificación de las técnicas es dividirlo en dos partes, el 70% para muestra de entrenamiento que consta de 1260 datos, y el 30 % para la muestra de validación y muestra de prueba que consta de 540 conjuntamente.

## **Tercer paso: Supuestos**

Análisis Discriminante Lineal de Fisher:

No se cumplieron los supuestos de Normalidad Multivariada y Homogeneidad de Varianzas, como suele ocurrir en la mayor parte de las aplicaciones prácticas. Al respecto Ruiz (2013) señala que “las suposiciones de normalidad e igualdad de varianzas no siempre se cumplen en las variables del modelo”. En el mismo sentido, Muñoz (1998) señala que “No obstante, el no cumplimiento de los supuestos de Normalidad Multivariada y Homogeneidad de Matriz Varianza Covarianza, no es problema para el análisis”.

Para verificar el supuesto de la colinealidad se calculó el Factor de Inflación de Varianza (FIV). En el Cuadro N° 4 el valor del Factor de Inflación de Varianza en cada curso es menor a 5, por lo que se puede concluir que no existe colinealidad.

Cuadro N° 4 Factor de Inflación de la Varianza

Aplicación 1:	
Curso	FIV
RM	1.785
RV	1.105
MAT	2.503
FIS	2.419
QUI	2.385
BIO	1.754

Fuente: Elaboración Propia

Máquina de Soporte Vectorial:

La técnica Máquina de Soporte Vectorial es más tolerante que el Análisis Discriminante Lineal de Fisher en la evaluación de los supuestos de Normalidad Multivariada y Homogeneidad de Varianzas. Esto, debido a que la técnica es de naturaleza no paramétrica y computacional.

Respecto al supuesto de la colinealidad se calculó el Factor de Inflación de Varianza (FIV). Según el cuadro N° 4 se observa que el factor de inflación de la varianza en cada curso es menor a 5, por lo que se puede concluir que no existe colinealidad.

#### **Cuarto Paso: Estimación de la regla de clasificación**

Análisis Discriminante Lineal de Fisher:

Para estimar la Función Discriminante Lineal de Fisher mediante el paquete estadístico R, se tomó los datos de la muestra de entrenamiento (70%) y se consideró el ingreso de todas las variables:

$$Z_i = -0.0236226RM - 0.047539RV - 0.1479MAT - 0.1012FIS + 0.0199QUI - 0.0871BIO$$

## Quinto Paso: Interpretación resultados

Análisis Discriminante Lineal de Fisher:

Según el cuadro N° 5, con un nivel de significancia del 1% se comprueba que todos los cursos discriminan adecuadamente el resultado de los postulantes en la prueba de admisión de la UNALM. Desde el punto de vista univariado, la variable que más discrimina es Física debido a que presenta un alto valor de F (918,88) y un valor bajo de lambda de Wilks (0,662): el curso que menos discrimina es Razonamiento Verbal por tener un bajo valor bajo de F (118,89) y un valor alto en el lambda de Wilks (0,938).

Cuadro N° 5. Anova Univariado

Curso	Lambda de Wilks	F	P.value
RM	0.798	453.805	0.000
RV	0.938	118.897	0.000
MAT	0.668	895.059	0.000
FIS	0.662	918.880	0.000
QUI	0.787	487.581	0.000
BIO	0.776	518.972	0.000

Fuente: elaboración propia

En el cuadro N° 6 se verifica que la Función Discriminante Lineal de Fisher es significativa a un nivel de significancia del 1%.

Cuadro N° 6. Significancia de la Función Discriminante

Función1	Lambda de Wilks	Chi-cuadrado	Sig.
Aplicación 1	0.565	1022.784	0.000

Fuente: Elaboración Propia

En el cuadro N° 7 se observa la correlación de cada curso con la Función Discriminante Lineal de Fisher encontrada. El curso de Física tiene una alta correlación (0,816), seguido de Matemática (0.805) y Biología (0.613). Los cursos Razonamiento Matemático (0.573) y Razonamiento Verbal (0.293) tienen los valores más bajos de correlación, lo que indica una baja capacidad discriminante.

Cuadro N° 7. Capacidad Discriminante de la Función

Curso	Función 1: Aplicación 1
FIS	0.816
MAT	0.805
BIO	0.613
QUI	0.594
RM	0.573
RV	0.293

Fuente: Elaboración Propia

Máquina de Soporte Vectorial:

Para obtener una óptima regla de clasificación se realiza un ajuste o modelamiento de los parámetros de la Máquina de Soporte Vectorial para cada Kernel. Previamente se fijan como parámetros iniciales los sugeridos por Hsu (2000),  $\text{Gamma} = 2^{(-1:1)}$  y  $\text{Costo} = 2^{(2:4)}$ . Luego, se afinan los parámetros establecidos y se determina los mejores parámetros mediante una Búsqueda Exhaustiva en Rejilla para cada Kernel (Lineal y Radial).

En el Cuadro N° 7 se observa que el mejor parámetro de Costo para el Kernel Lineal es 8 y el número de vectores de soporte es 75.

Cuadro N° 7. Regla de clasificación Kernel Lineal

```

SVM: Kernel Lineal
Parameter tuning of 'svm':
- sampling method: 10-fold cross
  validation
- best parameters: cost 8

Number of Support Vectors:75
    
```

Fuente: Elaboración Propia

El Cuadro N° 9 muestra que el mejor parámetro de Costo para el Kernel Radial es 2 y para Gama 0.25. El número de vectores de soporte para este caso es 169.

Cuadro N° 9. Regla de Clasificación Kernel Radial

<b>SVM: Kernel radial</b>	
Parameter tuning of 'svm':	
- sampling method: 10-fold cross validation	
- Best parameters: Cost gamma	
2	0.25
Number of Support Vectors: 169	
(99 70) Levels: 1 2	

Fuente: Elaboración Propia

### Sexto Paso: Validación de Resultados.

Análisis Discriminante Lineal de Fisher:

El Cuadro N° 10, muestra que la función clasificó correctamente al 96.83% de postulantes cuya preparación no realizó en el CEP-UNALM.

Cuadro N° 10 Validaciones de Resultado

Validación:		
	<b>SI Ingresó</b>	<b>NO Ingresó</b>
<b>SI Ingresó</b>	156	25
<b>NO Ingresó</b>	15	1064
Exactitud	: 0.9683	

Fuente: Elaboración propia

Máquina de Soporte Vectorial:

En el Cuadro N° 11 y Cuadro N° 12 se puede observar que el mejor función Kernel para la regla de decisión de la Máquina de Soporte Vectorial es el Lineal debido a que ofrece la mejor tasa de clasificación en la muestra de validación (97.93%), respecto al Kernel Radial (96.42)



Cuadro N° 11 Validaciones de Resultado

<b>SVM: Kernel Lineal</b>
10-fold cross-validation on training data: Total Accuracy: 97.93651 Single Accuracies:
97.61905
98.4127
97.61905
96.8254
96.03175
99.20635
98.4127
97.61905
99.20635

Cuadro N° 12 Validaciones de Resultado

<b>SVM: Kernel radial</b>
10-fold cross-validation on training data: Total Accuracy: 96.42857 Single Accuracies:
96.03175
95.2381
97.61905
97.61905
96.03175
94.44444
99.20635
96.8254
95.2381

### 6.1.3 Evaluación de las técnicas de clasificación

Para evaluar la clasificación de las técnicas Análisis Discriminante Lineal de Fisher y Máquina de Soporte Vectorial se tomó el 30% de los datos de prueba.

Análisis Discriminante Lineal de Fisher:

En el cuadro N° 13 los resultados altos de la diagonal de la matriz de Confusión 64 (si ingresó - si ingresó) y 456 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real.

Cuadro N° 13 Evaluación de Clasificación

Prediction	Aplicación 1	
	Si Ingresó	No Ingresó
Si Ingresó	64	5
No Ingresó	14	456

Fuente: elaboración Propia

En el Cuadro N° 14 se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (96,47%).

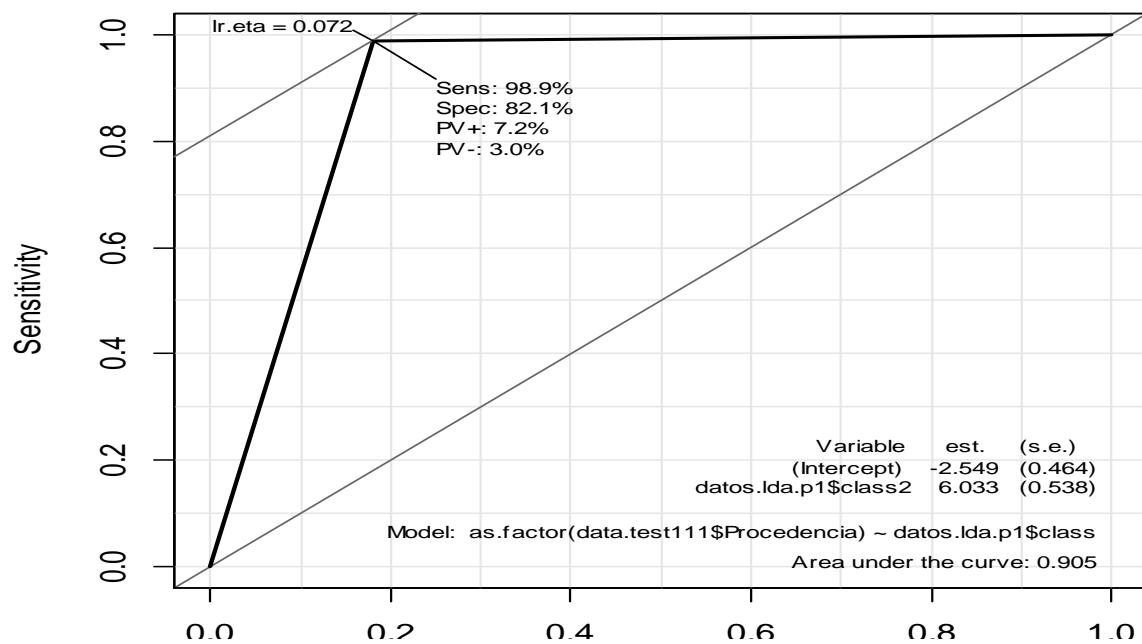
Cuadro N° 14 Evaluación de Clasificación

Statistics	Aplicación 1
Exactitud	0.9647
Sensibilidad	0.8205
Especificidad	0.9892
Valor pred (+)	0.9275
Valor pred (-)	0.9702

Fuente: elaboración Propia

En el gráfico N° 7 se puede observar que el área bajo la curva ROC es 0.905, este valor también reafirma la adecuada clasificación mediante el Análisis Discriminante Lineal de Fisher.

Gráfico N° 7. Área Bajo la Curva ROC



Fuente: Elaboración propia

## Máquina de Soporte Vectorial:

En el cuadro N° 15 los resultados altos de la diagonal de la matriz de Confusión 76 (si ingresó - si ingresó) y 450 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real cuando es empleado un Kernel lineal, por ello se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (97,59%).

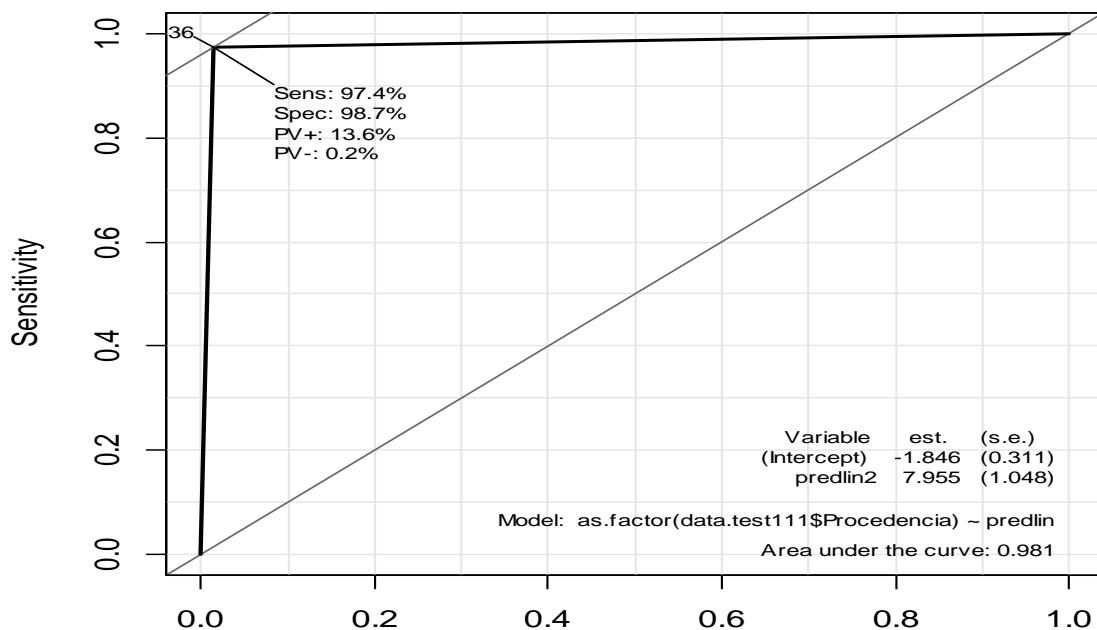
Cuadro N° 15 Evaluación clasificación Kernel Lineal

<b>SVM: Kernel lineal</b>		
Prediction	SI	NO
SI	76	12
NO	1	450
Exactitud	: 0.9759	
Sensibilidad	: 0.9870	
Especificidad	: 0.9740	
Valor pred (+)	: 0.8636	
Valor pred (-)	: 0.9978	

Fuente: Elaboración propia

En el gráfico N° 8 se puede observar que el área bajo la curva ROC es 0,981, este valor también reafirma la adecuada clasificación mediante La Máquina de Soporte Vectorial.

Gráfico N° 8 Área Bajo la Curva Roc Kernel Lineal



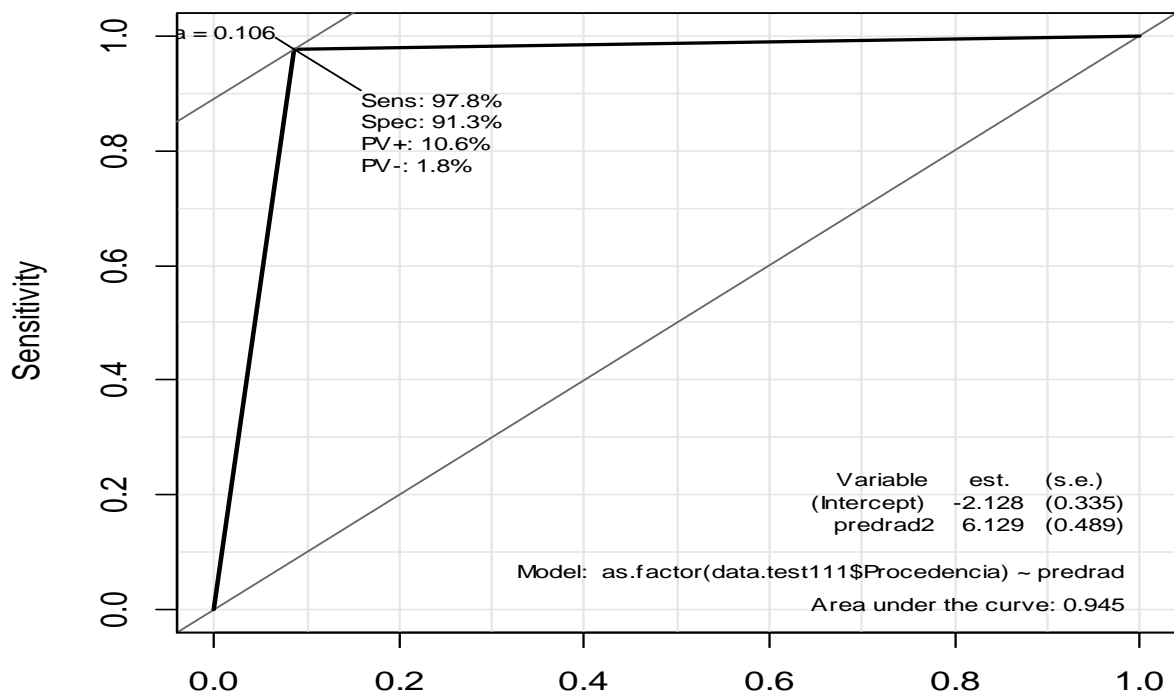
En el cuadro N° 16 los resultados altos de la diagonal de la matriz de Confusión 84 (si ingresó - si ingresó) y 437 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real cuando es empleado un Kernel Radial, por ello se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (96.66%).

Cuadro N° 16 Evaluación de Clasificación Kernel Radial

<b>SVM: Kernel radial</b>		
Prediction	SI	NO
SI	84	10
NO	8	437
Exactitud	: 0.9666	
Sensibilidad	: 0.9130	
Especificidad	: 0.9776	
Valor pred (+)	: 0.8936	
Valor pred (-)	: 0.9820	

En el gráfico N° 9 se puede observar que el área bajo la curva ROC es 0.945, este valor también reafirma la adecuada clasificación mediante la Máquina de Soporte Vectorial.

Gráfico N° 9 Área Bajo la Curva Roc Kernel Radial



De los resultados en cuanto al performance de clasificación para la regla de clasificación con la Máquina de Soporte Vectorial se ha determinado que el Kernel Lineal ofrece una mejor tasa de clasificación en la muestra de prueba (97,59%), respecto al Kernel Radial (96,42), así mismo tiene un Área Bajo la Curva Roc mejor de 0,981 vs 0,945 respectivamente.

## **6.2 Resultados aplicación 2**

### **6.2.1 Exploración de variables:**

El análisis descriptivo y gráfico de las variables independientes (puntajes de los cursos) en cada categoría de la variable dependiente (si ingresó, no ingresó) son los siguientes:

En el Cuadro N° 17 se muestra el puntaje promedio y desviación estándar de cada curso (razonamiento matemático, razonamiento verbal, matemáticas, física, química y biología) en cada condición de ingreso (si ingresó, no ingresó).

En el caso de los ingresantes el puntaje más alto se obtiene en el curso de Química (15.9022) y el más bajo en Razonamiento Verbal (8.6884). En el caso de los no ingresantes el puntaje más alto se registra en Química (9.5722) y el más bajo en Física (4.6807).

En el caso de los ingresantes la variabilidad más alta se presenta en el curso de Biología (3.67888) y el más bajo en Química (2.30370). En el caso de los no ingresantes la variabilidad más alta se presenta en el curso de Química (5.69752) y el más bajo en Razonamiento Verbal (3.02076).

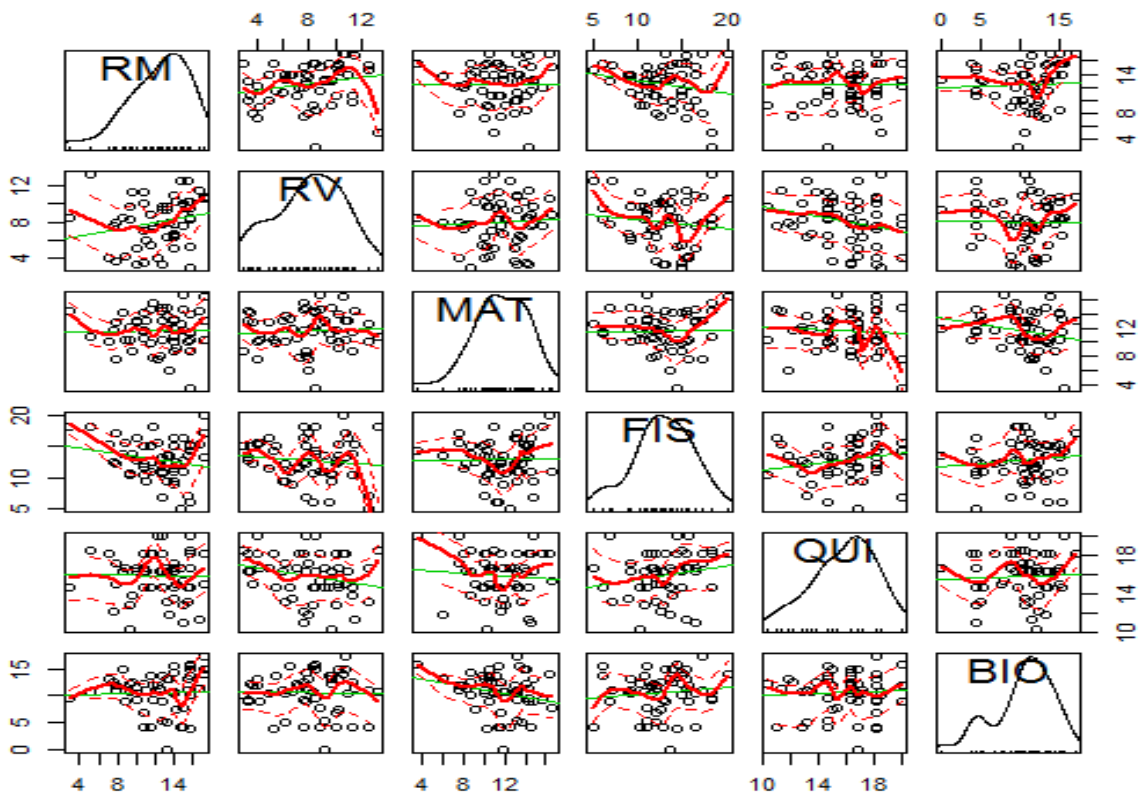
Cuadro N° 17: Resultados Descriptivos ingresantes y no ingresantes

CONDICIÓN	CURSO	MEAN	DESV
SI INGRESÓ	RM	12.4413	3.15260
	RV	8.0045	2.63281
	MAT	11.6153	2.64871
	FIS	12.8635	3.31251
	QUI	15.9022	2.30370
	BIO	10.5395	3.67888
NO INGRESÓ	RM	7.8519	4.33769
	RV	6.5870	3.02076
	MAT	5.5060	3.61936
	FIS	4.6807	4.42736
	QUI	9.5722	5.69752
	BIO	4.5027	4.16208

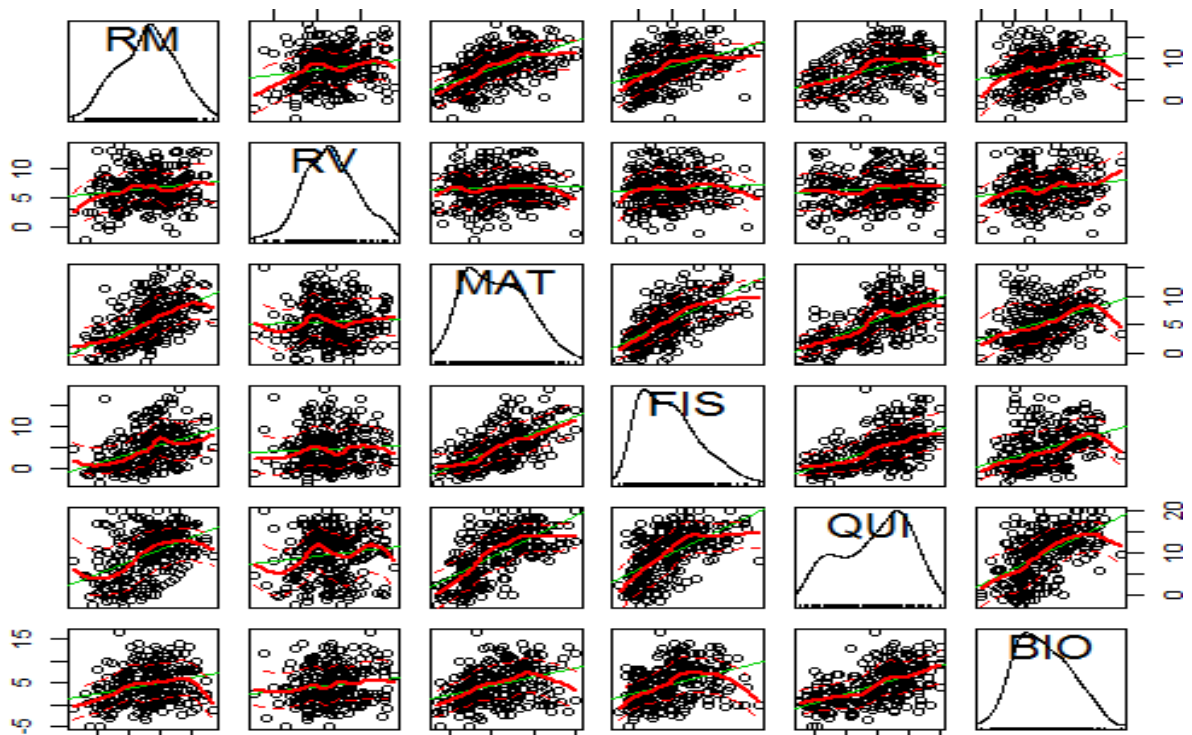
Fuente: Elaboración propia

Gráfico N° 10: Gráfico Descriptivo ingresantes y no ingresantes

Si Ingresantes

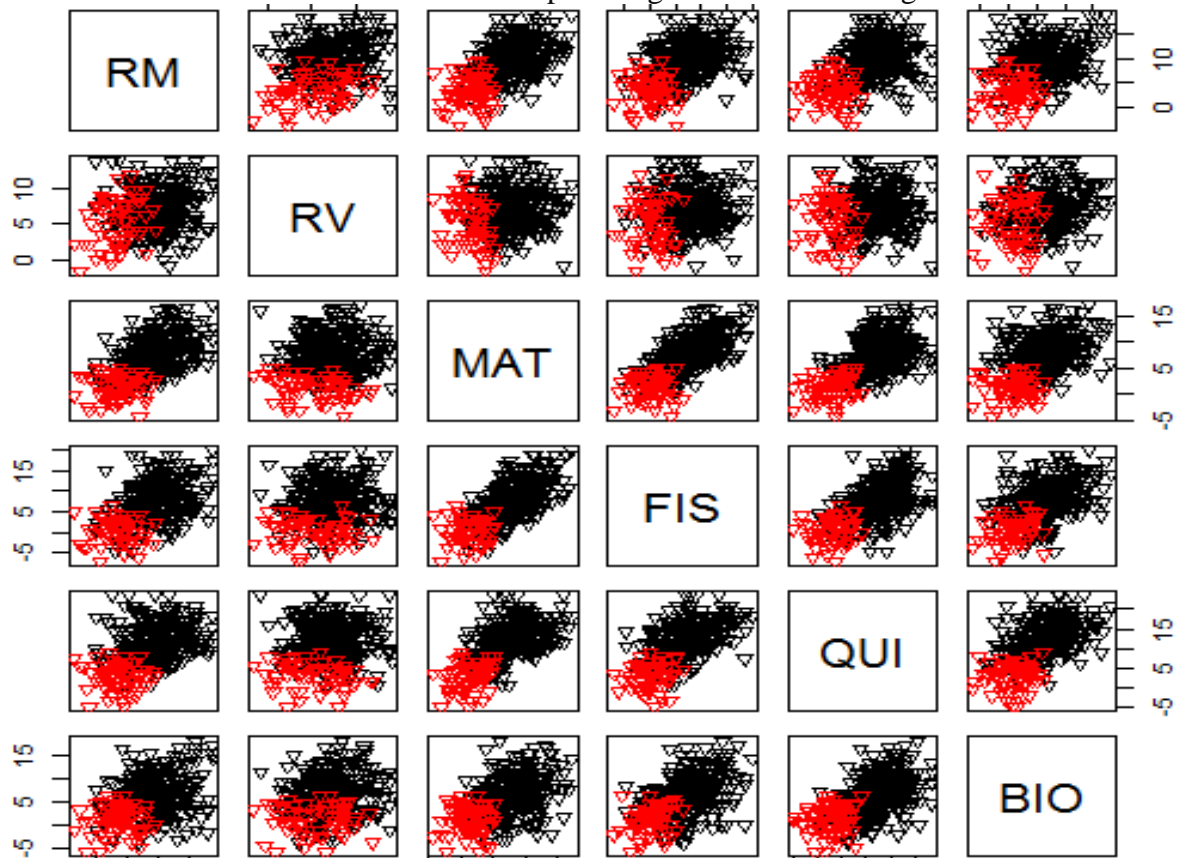


No Ingresantes



Fuente: Elaboración Propia

Grafico N° 11: Gráfico Descriptivo según condición de ingreso



Fuente: Elaboración propia

En el Grafico N° 10 se observa que no hay simetría en la distribución de las notas de los cursos, por lo que es previsible que no se cumpla el supuesto de normalidad univariada, el cual será verificado posteriormente. Además, se observa una tendencia lineal en la distribución de los cursos lo que permite sospechar la presencia de multicolinealidad, lo que también será verificado posteriormente.

En el Gráfico N° 11 se aprecia que los ingresantes (color rojo) y no ingresantes (color negro) están visiblemente diferenciados en la gráfica de dispersión, por lo que se supone que estarán divididos por un hiperplano lineal, lo cual se comprobará más adelante.

### **6.2.2 Aplicación de las técnicas de clasificación**

#### **Primer paso: Objetivos**

Los objetivos para La técnicas Análisis Discriminante Lineal de Fisher y La Máquina de Soporte Vectorial son los mismos que se ha propuesto para la primera aplicación, tan solamente que ahora aplicado al grupo de postulantes que ingresaron o no a la UNALM y cuya preparación fue en la CEP-UNALM

#### **Segundo paso: Diseño de la Investigación**

Las variables fueron definidas anteriormente. El tamaño de muestra considerado para la aplicación de ambas técnicas es 329 postulantes. La estrategia de partición para evaluar el poder de clasificación de las técnicas es dividirlo en dos partes, el 70% para muestra de entrenamiento y validación que incluyen a un total 230 postulantes y el 30%, para la muestra de prueba que incluyen a un total de 99 postulantes.



### Tercer paso: Supuestos

Supuestos Análisis discriminante lineal de Fisher:

Al igual que en la primera aplicación, no se ha cumplido los supuestos de Normalidad Multivariada y Homogeneidad de Varianzas.

Para verificar el supuesto de la colinealidad se calculó el Factor de Inflación de Varianza (FIV). En el Cuadro N° 18 el valor del Factor de Inflación de Varianza en cada curso es menor a 5, por lo que se puede concluir que no existe colinealidad.

Cuadro N° 18. Factor de Inflación de la Varianza

Aplicación 2:	
Curso	FIV
RM	1.693
RV	1.085
MAT	2.780
FIS	2.455
QUI	2.220
BIO	1.684

Fuente: Elaboración Propia

Supuestos de la Máquina de Soporte Vectorial:

De la misma forma como se ha precisado en la aplicación 1: la evaluación de supuestos en la Máquina de Soporte Vectorial a diferencia del Análisis Discriminante Lineal de Fisher es bastante tolerante, sin embargo de igual manera se ha realizado el diagnóstico colinealidad, el mismo que se observa en el cuadro N° 18 donde el factor de inflación de la varianza es menor a 5 por lo que podemos concluir que no existe colinealidad entre las variables independientes.

#### Cuarto Paso: Estimación de la regla de clasificación

Análisis Discriminante Lineal de Fisher:

Para estimar la Función Discriminante Lineal de Fisher mediante el paquete estadístico R, se tomó los datos de la muestra de entrenamiento (70%) y se consideró el ingreso de todas las variables:

$$Z_i = -0.00463RM - 0.04347RV - 0.12829MAT - 0.13031FIS + 0.046262QUI - 0.103BIO$$

#### Quinto Paso: Interpretación resultados

Análisis Discriminante Lineal de Fisher:

Según el cuadro N° 19, con un nivel de significancia del 1% se comprueba que todos los cursos discriminan adecuadamente el resultado de los postulantes en la prueba de admisión de la UNALM. Desde el punto de vista univariado, la variable que más discrimina es Física debido a que presenta un alto valor de F (172,921) y un valor bajo de lambda de Wilks (0,653): el curso que menos discrimina es Razonamiento Verbal por tener un bajo valor bajo de F (59,095) y un valor alto en el lambda de Wilks (0,965).

Cuadro N° 19. Anova Univariado

Curso	Lambda de Wilks	F	Sig.
RM a	0.847	59.095	0.000
RV a	0.965	11.820	0.001
MAT	0.693	144.681	0.000
FIS	0.653	172.921	0.000
QUI	0.829	67.470	0.000
BIO	0.757	104.471	0.000

Fuente: Elaboración propia

En el cuadro N° 20 se verifica la Función Discriminante Lineal de Fisher es significativa a un nivel de significancia del 1%.

Cuadro N° 20. Significancia de la Función Discriminante

Función	Lambda de Wilks	Chi-cuadrado	Sig.
Aplicación 2	,576	177,951	,000

Fuente: Elaboración Propia

En el cuadro N° 21 se observa la correlación de cada curso con la Función Discriminante Lineal de Fisher encontrada. El curso de Física tiene una alta correlación (0.805), seguido de Matemática (0.777) y Biología (0.666). Los cursos Razonamiento Matemático (0.497) y Razonamiento Verbal (0.222) tienen los valores más bajos de correlación, lo que indica una baja capacidad discriminante.

Cuadro N° 21. Capacidad Discriminante de la Función

Curso	Aplicación 2
FIS	0.850
MAT	0.777
BIO	0.660
QUI	0.531
RM a	0.497
RV a	0.222

Fuente: Elaboración Propia

Máquina de Soporte Vectorial:

El procedimiento que se sigue para obtener la mejor regla de decisión es el mismo que se ha realizado en la aplicación1, es decir se procede a realizar un modelamiento de los parámetros de las Máquina de Soporte Vectorial para cada Kernel, pero antes como paso previo se fijan los parámetros iniciales  $\text{Gamma} = 2^{-1:1}$  y  $\text{Costo} = 2^{(2:4)}$ .

En el Cuadro N° 22 se observa que el mejor parámetro de Costo para el Kernel Lineal es 8 y el número de vectores de soporte es 29.

Cuadro N°22. Regla de Clasificación Kernel lineal

```
SVM: Kernel lineal  
Parameter tuning of 'svm':  
- sampling method: 10-fold cross  
validation  
- best parameters: cost=8  
Number of Support Vectors: 29 (16 13 )
```

Fuente: Elaboración Propia

En el Cuadro N° 23 se observa que el mejor parámetro de Costo para el Kernel Radial Lineal es 2 y para el parámetro Gama es 0.125. El número de vectores de soporte para este caso es 48.

Cuadro N°23. Regla de Clasificación Kernel radial

```
SVM: Kernel Radial  
Parameter tuning of 'svm':  
- sampling method: 10-fold cross  
validation  
- best parameters:  
cost gamma  
2 0.125  
Number of Support Vectors: 48 ( 26 22 )
```

Fuente: Elaboración Propia

### **Sexto paso: Validación de Resultados.**

Análisis Discriminante Lineal de Fisher:

En el Cuadro N° 24, la función Análisis Discriminante Lineal de Fisher clasificó correctamente en un 93.48% a los postulantes cuya preparación no realizó en el CEP-UNAL en sus dos condiciones de ingreso.

Cuadro N° 24 Validaciones de Resultado

		Validación:	
		SI Ingresó	NO Ingresó
SI Ingresó		30	7
NO Ingresó		8	185
Exactitud		: 0.9348	

Máquina de Soporte Vectorial:

En el Cuadro N° 25 y Cuadro N° 26 se puede observar que el mejor Kernel para la Máquina de Soporte Vectorial es el Lineal debido a que ofrece una mejor tasa de clasificación en la muestra de validación (94,78%), respecto al Kernel Radial (95,21)

Cuadro N° 25 Validaciones de Resultado

<b>SVM: Kernel lineal</b>
10-fold cross-validation on training data:
Total Accuracy: 94.78261
Single Accuracies:
100
95.65217
100
91.30435
95.65217
95.65217
86.95652
95.65217
91.30435
95.65217

Cuadro N° 26 Validaciones de Resultado

<b>SVM: Kernel Radial</b>
10-fold cross-validation on training data:
Total Accuracy: 95.21739
Single Accuracies:
91.30435
95.65217
86.95652
86.95652
91.30435
95.65217
91.30435
100
95.65217

Fuente: Elaboración Propia

### 6.2.3 Evaluación de las técnicas de clasificación.

Para evaluar la clasificación de las técnicas Análisis Discriminante Lineal de Fisher y Máquina de Soporte Vectorial se tomó el 30% del total de datos para cada uno, lo mismo que se ha llamado muestra de validación y de prueba.

## Análisis Discriminante Lineal de Fisher:

En el cuadro N° 27 los resultados altos de la diagonal de la matriz de Confusión 10 (si ingresó - si ingresó) y 82 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real.

Cuadro N° 27 Evaluación de la Clasificación

	2da Aplicación(SI CEP)	
Prediction	Si Ingresó	No Ingresó
Si Ingresó	10	3
No Ingresó	3	82

Fuente: Elaboración Propia

En el Cuadro N° 28 se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (93,88%).

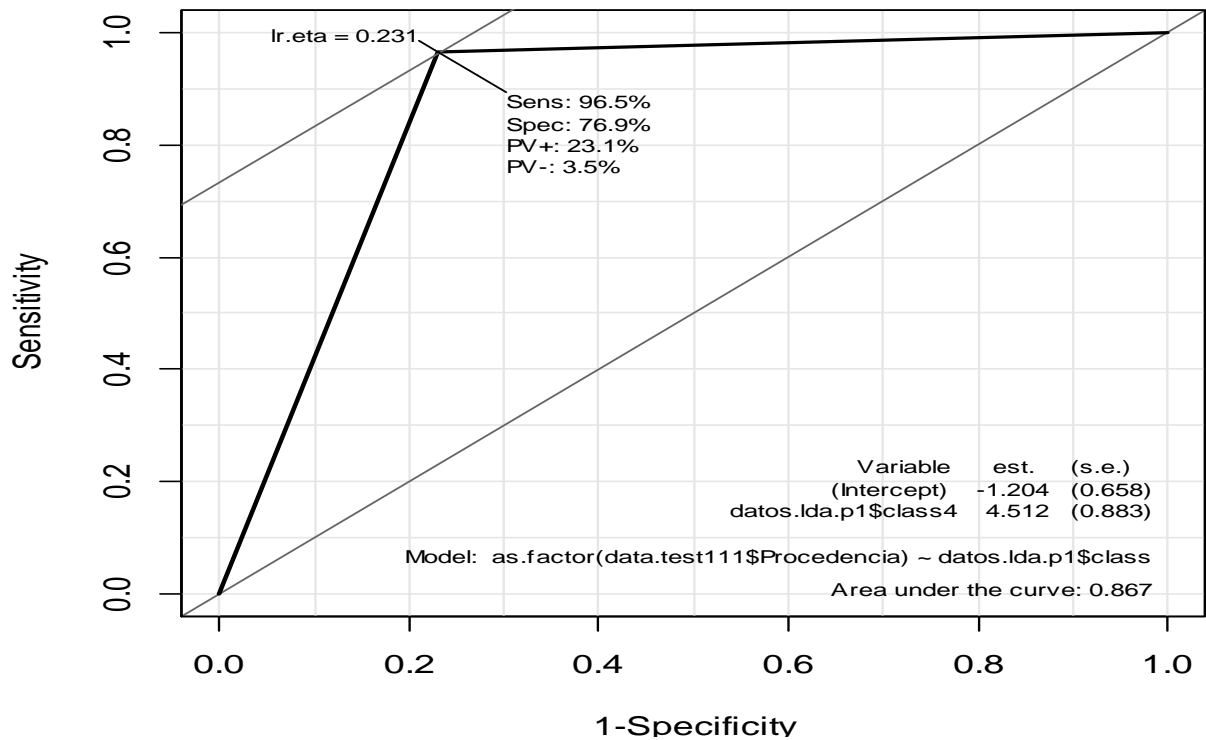
Cuadro N° 28 Evaluación de la Clasificación

Statistics	Aplicación 2
Exactitud	0.9388
Sensibilidad	0.7692
Especificidad	0.9647
Valor pred (+)	0.7692
Valor pred (-)	0.9647

Fuente: elaboración Propia

En el gráfico N° 12 se puede observar que el área bajo la curva ROC es 0.867, este valor también reafirma la adecuada clasificación mediante el Análisis Discriminante Lineal de Fisher.

Gráfico N°12. Área Bajo la Curva Roc



Fuente: Elaboración Propia

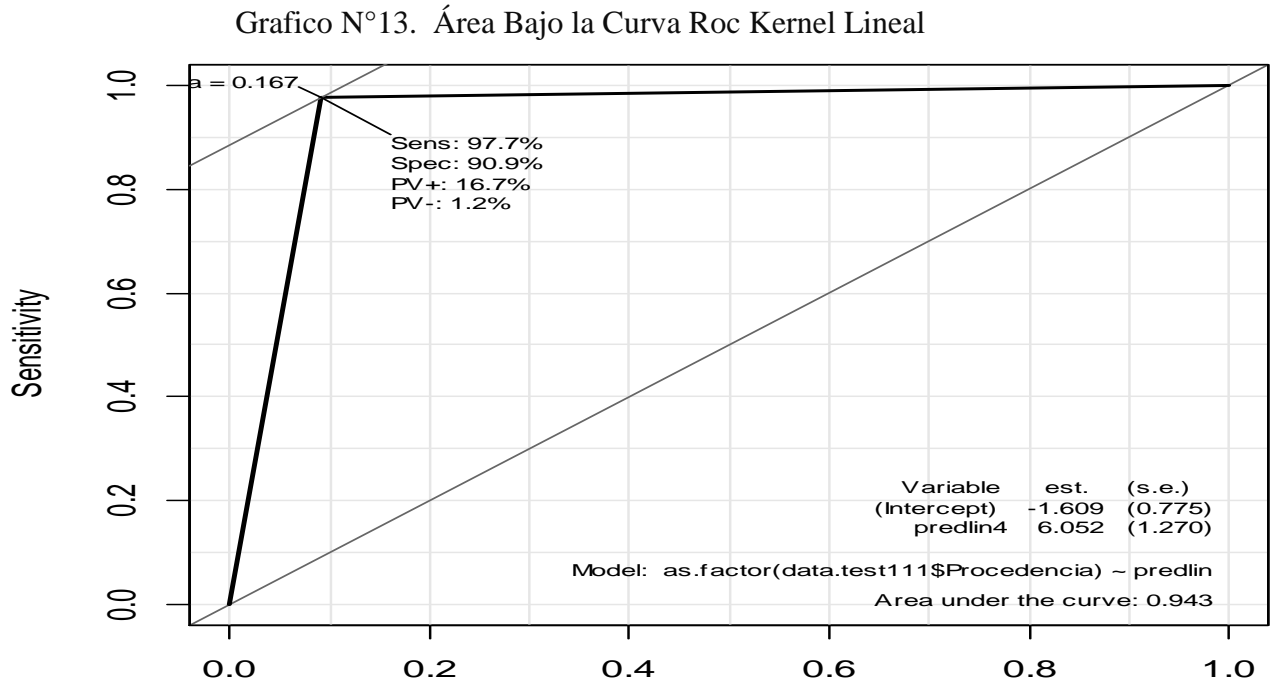
Máquina de Soporte Vectorial:

En el cuadro N° 29 los resultados altos de la diagonal de la matriz de Confusión 10 (si ingresó - si ingresó) y 85 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real cuando es empleado un Kernel lineal, por ello se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (96.94%).

Cuadro N°29 Evaluación Clasificación Kernel Lineal

<b>SVM: Kernel Lineal</b>		
Prediction	SI	NO
SI	10	2
NO	1	85
Exactitud	:	0.9694
Sensibilidad	:	0.9091
Especificidad	:	0.9770
Valor pred (+)	:	0.8333
Valor pred (-)	:	0.9884

En el gráfico N° 13 se puede observar que el área bajo la curva ROC es 0,943, este valor también reafirma la adecuada clasificación mediante La Máquina de Soporte Vectorial.



Fuente: Elaboración Propia

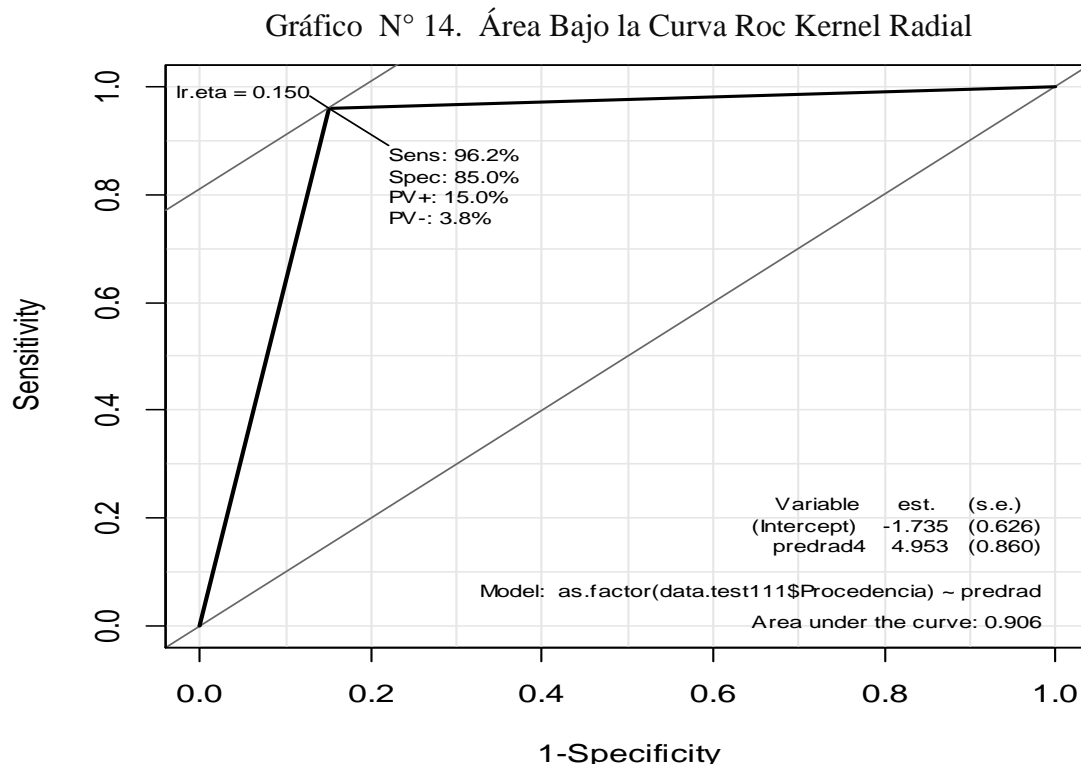
En el cuadro N° 30 los resultados altos de la diagonal de la matriz de Confusión 17 (si ingresó - si ingresó) y 75 (no ingresó - no ingresó) permiten afirmar que es buena la correspondencia entre el modelo predictivo y el modelo real cuando es empleado un Kernel Radial, por ello se puede observar que todos los indicadores de clasificación presentan un buen performance de clasificación, siendo la Exactitud o Tasa de Buena Clasificación uno de los indicadores más importantes (93.88%).

Cuadro N° 30 Evaluación Clasificación Kernel Radial

<b>SVM: Kernel radial</b>		
predrad	SI	NO
SI	17	3
NO	3	75
Exactitud	: 0.9388	
Sensibilidad	: 0.8500	
Especificidad	: 0.9615	
Valor pred (+)	: 0.8500	
Valor pred (-)	: 0.9615	



En el gráfico N°14 se puede observar que el área bajo la curva ROC es 0.906, este valor también reafirma la adecuada clasificación mediante la Máquina de Soporte Vectorial.



Fuente: Elaboración Propia

De los resultados en cuanto al performance de clasificación para la Máquina de Soporte Vectorial se ha determinado que el Kernel Lineal ofrece una mejor tasa de clasificación en la muestra de prueba (96,94%), respecto al Kernel Radial (93,88%), así mismo tiene un Área Bajo la Curva Roc mejor de 0,943 vs 0,905 respectivamente.

### 6.3. Comparación de Resultados:

Cuadro N°31. Comparación de las clasificaciones

#### 1era Aplicación- No CEP-UNALM

	Precisión	Sensibilidad	Especificidad	PV+	PV-	AUC
Función Discriminante	0.9647	0.8205	0.9892	0.9275	0.9702	0.905
Regla SVM lineal	0.9759	0.9870	0.9740	0.8636	0.9978	0.981

#### 2da Aplicación-SI-CEP-UNALM

	Precisión	Sensibilidad	Especificidad	PV+	PV-	AUC
Función Discriminante	0.9388	0.7692	0.9647	0.7692	0.9647	0.8607
Regla SVM lineal	0.9694	0.9091	0.9770	0.833	0.988	0.943

Fuente: Elaboración Propia

Según el cuadro N° 31 se observa una mejor tasa de buena clasificación con la técnica Máquina de Soporte Vectorial en el resultado en la prueba de admisión de la UNALM de los postulantes que recibieron o no preparación en el CEP-UNALM, de la misma forma se observa una mayor Área Bajo la Curva ROC.

Debido a que en ambas aplicaciones se observa una similar distribución en la clasificación para cada técnica respecto a cada aplicación y sumado a ello, gracias a que los resultados del Análisis Discriminante permitieron perfilar a los postulantes con los mismos cursos en función de la procedencia se puede concluir que no existe diferencia en perfil de los alumnos.

## VII. CONCLUSIONES

1. La técnica Máquina de Soporte Vectorial clasifica mejor el rendimiento en la prueba de admisión de la UNALM de los postulantes que recibieron o no preparación en el CEP-UNALM, al presentar tasas de buena clasificación de 96.94% y 97.59% respectivamente, en relación a la técnica de Análisis Discriminante Lineal de Fisher de 93.88% y 96.47%.
2. En la técnica Máquina de Soporte Vectorial, el Kernel de tipo Lineal clasifica mejor el rendimiento en la prueba de admisión de la UNALM de los postulantes que recibieron o no preparación en el CEP-UNALM, al presentar el área bajo la curva ROC de 0.94 y 0.981 respectivamente, en relación al kernel de tipo radial 0.906 y 0.94
3. En la técnica Máquina de Soporte Vectorial, el Kernel de tipo Lineal clasifica mejor el rendimiento en la prueba de admisión de la UNALM de los postulantes que recibieron o no preparación en el CEP-UNALM, al presentar tasas de buena clasificación de 96.94% y 97.59% respectivamente, en relación al kernel de tipo radial de 93.88% y 96.66 %.
4. La técnica Máquina de Soporte Vectorial clasifica mejor el rendimiento en la prueba de admisión de la UNALM de los postulantes que recibieron o no preparación en el CEP-UNALM, al presentar el área bajo la curva ROC de 0.943 y 0.981 respectivamente, en relación a la técnica de Análisis Discriminante Lineal de Fisher de 0.8607 y 0.905
5. En los postulantes que no recibieron preparación en el CEP-UNALM, el Análisis Discriminante Lineal de Fisher los cursos Física (0,816), Matemática (0,805) y Biología (0,613) discriminan mejor el rendimiento en la prueba de admisión, a diferencia de Razonamiento Verbal (0,573) y Razonamiento Matemático (0,293) que poseen la menor capacidad discriminante.
6. En los postulantes que si recibieron preparación en el CEP-UNALM, el Análisis Discriminante Lineal de Fisher los cursos Física (0,859), Matemática (0,777) y Biología (0,66) discriminan mejor el rendimiento en la prueba de admisión, a diferencia de Razonamiento Verbal (0,497) y Razonamiento Matemático (0,222) que poseen la menor capacidad discriminante.

7. Se puede concluir en base a las dos conclusiones precedentes que la formación que brinda actualmente en la CEP-UNALM, no proporciona a los postulantes que se prepararon en ella un rendimiento diferenciado en la prueba de admisión de la universidad con respecto a los postulantes que se prepararon en otras academias

## VIII. RECOMENDACIONES

1. Comparar la capacidad de clasificación de la Máquina de Soporte Vectorial y el Análisis Discriminante Cuadrático, ante el incumplimiento del supuesto de homogeneidad de varianza en esta última técnica.
2. Comparar la capacidad de clasificación de la Máquina de Soporte Vectorial y el Análisis Discriminante Lineal de Fisher cuando la variable dependiente tiene más de dos clases.
8. Realizar un estudio para identificar las razones que determinan que en la actualidad la formación que brinda actualmente en la CEP-UNALM, no proporciona a los postulantes que se prepararon en ella un rendimiento diferenciado en la prueba de admisión de la universidad con respecto a los postulantes que se prepararon en otras academias

#### IV REFERENCIAS BIBLIOGRÁFICAS

Abril, G., (2003). Modelos de Clasificación basados en la Máquina de Vectores soporte. Departamento de Economía Aplicada I. Universidad de Sevilla.

Ali, S. and Smith, K., (2006) A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing*, 70(1-3):173-186.

Alonso, R. Talbot, M. (1992) Inteligencia Artificial y estadística. *Estadística Española*, Vol 34 (131), p. 407-430.

Aluja T. (2001) La minería de datos, entre la estadística y la inteligencia artificial. Universidad de Valencia. vol. 25(3), p. 479-498,

Perez, C. (2007), Minería de datos. Técnicas y Herramientas, Thompson, Madrid-España.

Chávez, M., Casas, G., Moerira, J., Bello, R., Grau, R., (2009). Perfeccionamiento de la matriz de confusión que resulta de un clasificador, en dependencia del dominio de aplicación. Centro de Estudios de Informática, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Villa Clara, Cuba.

Cristianini, N., Shawe, J., Taylor, J., (2000) An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 190p.

Garcia, E., Lozano, F., (2006) Boosting Support Vector Machines, Bogotá-Colombia, Disponible en: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0121-49932006000200008](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-49932006000200008)

Guzman, M.V. Carrillo H. Villaseñor, E. Calero,R. Morán,L.E. Acosta, A. (2004), Minería de Datos con Redes Neuronales Artificiales: Aplicación en Vacunas – Tuberculosis, X Convención Internacional y Feria Informática 2004, I Congreso de Bioinformática. Palacio de las Convenciones, La Habana, mayo del 2004.

Hamel, L., (2009) Knowledge Discovery With Support Vector Machine. Wiley, Canada .247p.

Hsu,W., Chang, C., Lin, C., (2003) A practical guide to support vector classification. Department of Computer Science National Taiwan University, Taipei 106, Taiwan.

Hair, Anderson y Tatham, Black (1999). Análisis Multivariante. Madrid: Prentice Hall Iberia.

Hastie, T; Tibshirani, R y Friedman, J. 2008. The elements of statistical learning.2 ed. SPRINGER.

International conference 4th, MLDM 2005 leipzig, germany, july 2005: Machine learning in pattern recognition,Proceedings Springer 695p.

James, G., Witten, D. Hastie, T., Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*, Springer Verlag, 2013

Joachims, T.,(2001) Learning To Classify Text Using Support Vector Machines. Springer, New York. 205p.

Lee, J. B., Park, M. & Song, H. S. (2005), 'An extensive comparison of recent classification tools applied to microarray data', Computational Statistics & Data Analysis 48, p. 869-885

Levy, M., Varela, M. (2003) Análisis multivariable para las ciencias sociales, Pearson Education, Madrid-España.

Katz, M., (2011) Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers, Third Edition. Cambridge University Press. University of California, USA. 233p.

Maldonado, S., Weber, R., (2012) Modelos de Selección de Atributos para Support Vector Machines. Revista Ingenieria de Sistema. 26:40-70

Martinez, R., (2008). Introducción a los métodos Kernel. Universidad Autonoma de Madrid Disponible en [http://arantxa.ii.uam.es/~jms/seminarios\\_doctorado/abstracts2007-2008/20080429MMartinez.pdf](http://arantxa.ii.uam.es/~jms/seminarios_doctorado/abstracts2007-2008/20080429MMartinez.pdf).

Mohammed, J.,Wagner, M., (2014). Data Mining and Analysis: fundamental and algorithms. Cambridge University Press, New York. 583p.

Montanero, J. 2008. Análisis multivariante. Universidad de Extremadura.

Morales, F. (2009) Aprendizaje con Kernels.México.(Consultado en 20 febrero 2015) disponible en <http://ccc.inaoep.mx/~emorales/Cursos/>.

Muñoz, S., (1998) La técnica de análisis discriminante: una aplicación para el área bancaria1, Costa Rica, Disponible en: [http://www.bccr.fi.cr/investigacioneseconomicas/sistemafinanciero/La\\_tecnica\\_analisis\\_discriminante\\_aplicacion\\_area\\_bancaria.pdf](http://www.bccr.fi.cr/investigacioneseconomicas/sistemafinanciero/La_tecnica_analisis_discriminante_aplicacion_area_bancaria.pdf)



Osuna, E., Freund, R., Girosit, F.,(1997) Training support vector machines: an application to face detection. In Computer Vision and Pattern Recognition. Proceedings., 1997 IEEE Computer Society Conference on. 1997.

Parr, R.,(2001) Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management, John Wiley & Sons – New York

Pitarque,A. Roy, J.F. Ruiz,J.C. (1998). Redes neuronales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación (1998) 19, 387-400

Platt, J., (1999) Fast training of support vector machines using sequential minimal optimization. MIT Press: Cambridge. pp 185-208.

Shigeo, A.,(2005) Support Vector Machine For Pattern Classification: Advances in pattern Recognition. Springer, USA. 343p.

Siddiqi, N (2012) Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring, Wiley and Sons, New York.

Tejeda, H. J (2004), Los modelos conexionistas y sus potencialidades en el análisis de datos 2 (1), p.11-20

Uriel , E., y Aldas, J, 2005. Análisis Multivariante Aplicado. 1 ed. THOMSON.

Vapnik, V., (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY.

Velásquez, J., Olaya, Y., Franco, C. (2010). Time Series Prediction Using Support Vector Machines. Revista Chilena de Ingenieria. 8(1). pp 64-75.

Vera, H. (2003), Evaluación de errores de datos sísmológicos y sus efectos sobre los parámetros y modelos derivados de éstos para una región determinada cubierta por la red del Observatorio Sísmológico del SurOccidente – OSSO ( Tesis Pos grado) Universidad del Valle. Colombia

Wang. L., 2005. Support Vector Machine: Theory and applicattions.Springer, Singapur. 431p.

Webb, A., Copsey, K., (2002) Statistical Pattern Recognition, Second Edition.Wiley,UK. 496p.

Wu, X., Kumar, V., (2009)The Top Ten Algorithms in Data Mining. Chapman Hall,USA.201p.

## X. ANEXOS:

Códigos en R:

```
library(e1071)
```

```
library(caret)
```

```
library(Epi)
```

```
Library(Klar)
```

Análisis Discriminante lineal de Fisher Primera Aplicación

```
data111<-read.table(file.choose(),T)
```

```
data.indices111 <- sample(1:nrow(data111),size=1260)
```

```
data.entrenamiento111 <- data111[data.indices111,]
```

```
data.test111 <- data111[-data.indices111,]
```

```
nrow(data.entrenamiento111)
```

```
library(MASS)
```

```
mod31<-lda(Procedencia~.,data=data.entrenamiento111 )
```

```
datos.lda.p1 <- predict(mod31, newdata=data.test111[,1:6])
```

```
mc11 <- table(datos.lda.p1$class, data.test111$Procedencia)
```

```
mc11
```

```
library(caret)
```

```
confusionMatrix(datos.lda.p1$class, data.test111[,7])
```

```
library(Epi)
```

```
ROC(form=as.factor(data.test111$Procedencia)~datos.lda.p1$class,
```

```
plot="ROC" )
```

Máquina Soporte Vectorial kernel Lineal

```
data111<-read.table(file.choose(),T)
```

```
data.indices111 <- sample(1:nrow(data111),size=1260)
```

```

data.entrenamiento111 <- data111[data.indices111,]
data.test111 <- data111[-data.indices111,]
nrow(data.entrenamiento111)
x<-data.entrenamiento111[1:6]
y<-data.entrenamiento111[7]

head(data.test111)

tune.outli<- tune(svm,train.x=x,train.y=y,kernel="linear",
ranges=list(cost=2^(2:4)),scale=T)
summary(tune.outli)
ccli <- as.numeric(tune.outli$best.parameters[1])
modelolin <- svm(as.factor(Procedencia)~., data.entrenamiento111[,1:7], type
="C-classification", cost = ccli, kernel = "linear",scale=T,cross=10)
summary(modelolin)
predlin <- predict(modelolin,data.test111[,1:6]);
confusionMatrix(predlin, data.test111[,7])
nrow(y)
nrow(predlin)
ROC( form=as.factor(data.test111$Procedencia)~predlin,plot="ROC")

```