

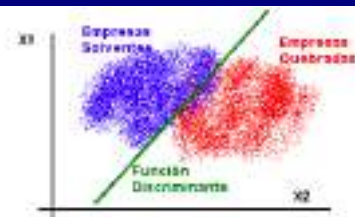


UNIVERSIDAD NACIONAL AGRARIA LA MOLINA
Dpto. académico de Estadística e Informática

Análisis Discriminante



Mg. Jesús Salinas Flores



jsalinas@lamolina.edu.pe

Reconocimiento de Patrones



Mg. Jesús Salinas Flores



jsalinas@lamolina.edu.pe



Reconocimiento de Imágenes

a a a *a* a a a a a a a

e e e *e* e e e e e e e

u u u *u* u u u u u u u



Arqueología Cuantitativa









1200 ac – 200 dc

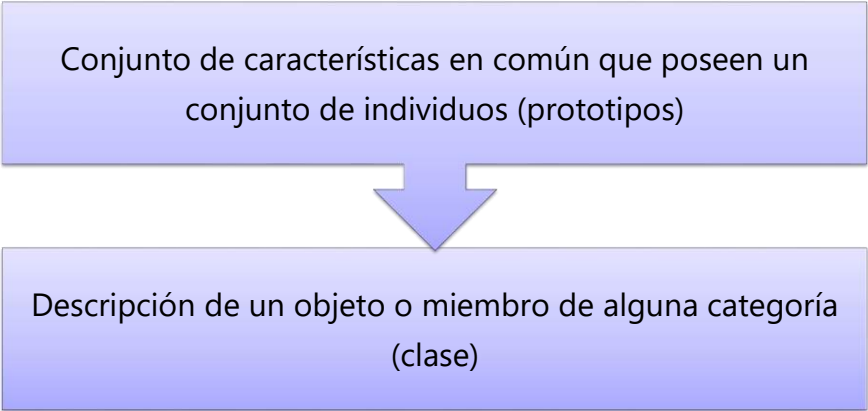
200 ac – 700 dc

200 dc – 700 dc

800 dc – 1400 dc



¿Qué es un patrón?

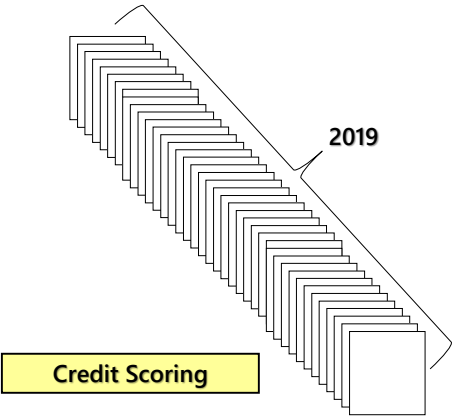


Morosidad en los créditos de consumo

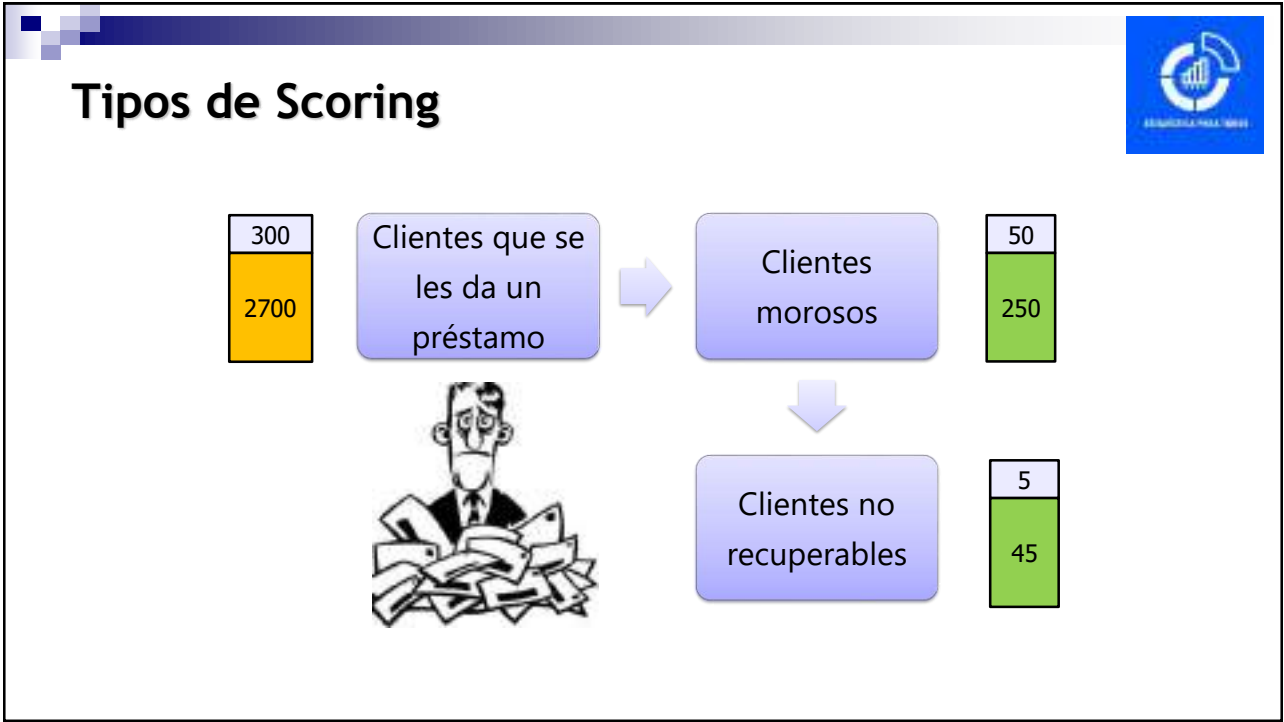
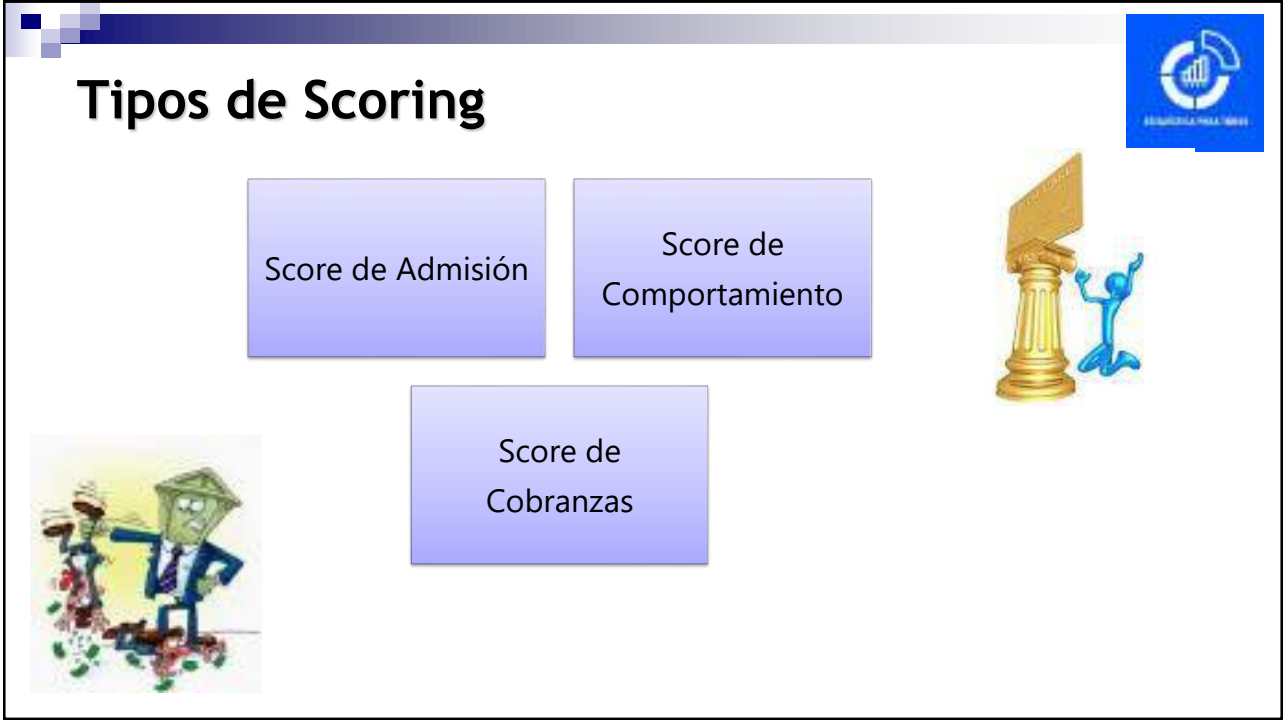
No Morosos: buenos pagadores (96%)



Morosos: malos pagadores (4%)



| | |
|---|-------|
| Ficha de Solicitud | |
| Género: | |
| Edad: | |
| Estado civil: | |
| Carga Familiar: | |
| Distrito: | |
| Monto del préstamo: | |
| Nivel de educación: | |
| Tipo de ocupación: | |
| Tipo de renta: | |
| Poseción de tarjeta de crédito: | |
| Poseción de propiedades: | |
| Tiempo de permanencia en el domicilio actual: | |
| Historial como aval: | |






Tipos de Scoring

Score de Admisión

Score de Comportamiento

Score de Cobranzas





Rendimiento académico en un semestre

Agosto



Diciembre



Alto (14-20)

Regular (11-13)

Bajo (0-10)

Promedio Ponderado Semestral



Número de cursos matriculados

Promedio semestral anterior

Promedio semestral acumulado

Número de créditos aprobados

Carrera

Edad


Género

Nivel Socio Económico




Situación académica al finalizar un curso


Agosto




Aprobado (11-20)




Desaprobado (0-10)



Diciembre



Nota final del curso



Número de cursos matriculados

Promedio semestral anterior

Promedio semestral acumulado


Nota del curso pre-requisito

Número de veces que lleva el curso

Edad

Género

Nivel Socio Económico





Policía de Richmond


Optimización y despliegue de recursos

Despliegue de fuerzas para

- Prevenir crímenes
- Determinar si una amenaza es real o no

Disminución de crímenes y homicidios entre el 20 y 30 % en un año





Universidad de Londres



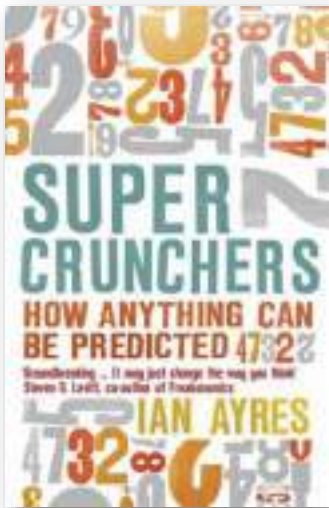
Necesidad

- Analizar el rendimiento académico de los estudiantes
- Anticipar el rendimiento académico de sus estudiantes






Beneficios

- Hoy en día se pueden identificar de manera **anticipada** los estudiantes que tendrán desempeño **exitoso** al momento de su ingreso.



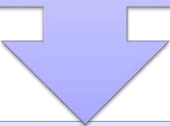
- ❖ Los "supercalculadores" son personas que están implementando esta nueva forma de **pensamiento numérico** en diversas áreas.
- ❖ El libro habla de la aplicación del reconocimiento estadístico de patrones a temas de todos los días: negocios, medicina, educación, etc.







- ❖ Muestra las últimas novedades en Big Data
- ❖ El autor comparte puntos de vista sobre la manera en la que diversas empresas están usando Big Data y la analítica de datos para resolver problemas del mundo real.

Reelección de Obama

Un equipo de estadísticos y especialistas de la información recolectaron los datos de millones de personas, permitiéndoles predecir y hasta impulsar a los votantes a que apoyen a su candidato.



Este equipo fue liderado por el científico **Rayid Ghani**, especialista en maximizar la eficiencia de las ofertas en los supermercados.



Estadística Criminal



Conceptos Básicos de un SRP

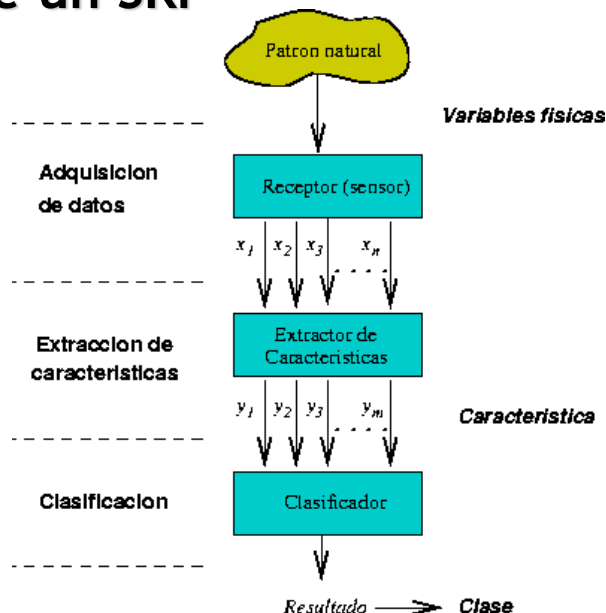


Clase - Categoría determinada por algunos atributos comunes a sus miembros.

Vector de características - Contiene la medida de las características de un patrón

Patrón - Descripción de los miembros de alguna categoría.

Esquema de un SRP



Variables Predictoras

Variables de Identificación

- Nombres, apellidos, correo, teléfono

Variables Socio-Demográficas

- Edad, Género, Ubigeo, NSE, Nivel de educación, Ocupación, Número de dependientes, etc.

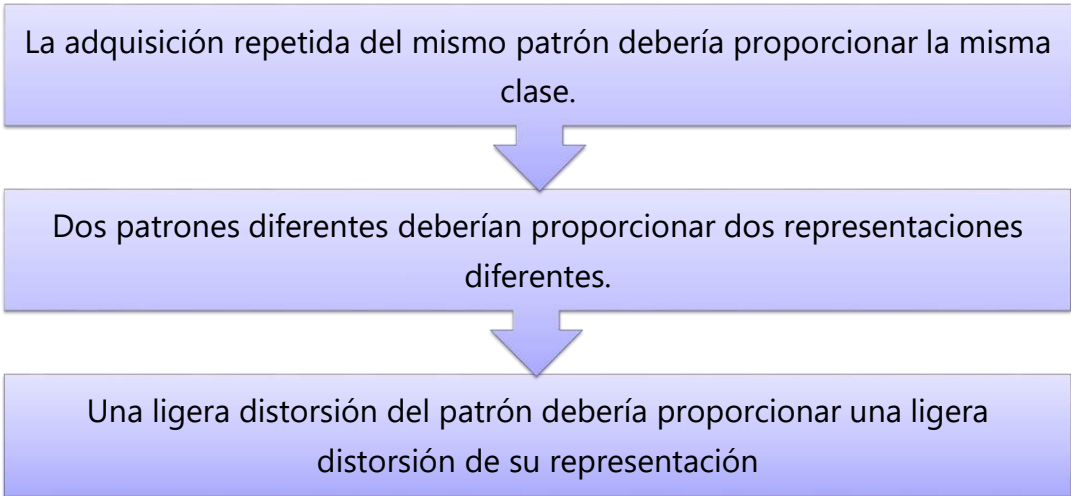
Variables del Negocio (Comportamiento)

- Expresan la información de cada cliente con el negocio: monto del préstamo, días de mora, número de llamadas perdidas, plan adquirido, número de reclamos, etc.

Variables de Percepción del Cliente (Actitudinal)

- Encuesta de satisfacción o de actitud hacia un servicio o producto.

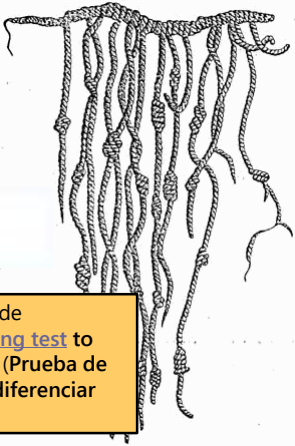
SRP perfecto



Cultura Paracas



700 ac- 200 dc



Cultura Cupisnique

Captcha o CAPTCHA son las siglas de Completely Automated Public Turing test to tell Computers and Humans Apart (Prueba de Turing pública y automática para diferenciar máquinas y humanos).^[1]

Locke

Radicati



Aprendizaje y/o Entrenamiento



Se suele utilizar los términos **aprendizaje** y **entrenamiento** para referirse al proceso de construcción del clasificador.

Aprendizaje supervisado

Aprendizaje no supervisado

Tipos de Aprendizaje



Aprendizaje Supervisado

- Se tiene un conjunto de prototipos de los cuales se conoce su clase. A este conjunto se le denomina **conjunto de entrenamiento**. Las técnicas utilizadas son las **técnicas de regresión** y las de **clasificación o discriminantes**.

Aprendizaje No Supervisado

- Se tiene un conjunto de prototipos del que no se conoce su clase cierta. En ocasiones, ni siquiera se conoce el número de clases. Las técnicas utilizadas suelen denominarse métodos de agrupamiento o **clustering**.

Técnicas Discriminantes



Mg. Jesús Salinas Flores



jsalinas@lamolina.edu.pe

Introducción

Las TD se utilizan para clasificar a distintos individuos en grupos o poblaciones ya conocidos a partir de los valores de un conjunto de variables sobre los individuos a los que se pretende clasificar.



$$\text{Modelo: } Y = f(X_1, X_2, \dots, X_p)$$



La variable dependiente Y (categórica) se denomina VARIABLE CRITERIO, TARGET y las variables X_1, X_2, \dots, X_p (independientes) se denominan VARIABLES PREDICTORAS o clasificadoras y pueden ser cuantitativas y/o cualitativas





Finalidad de las TD

Explicativos

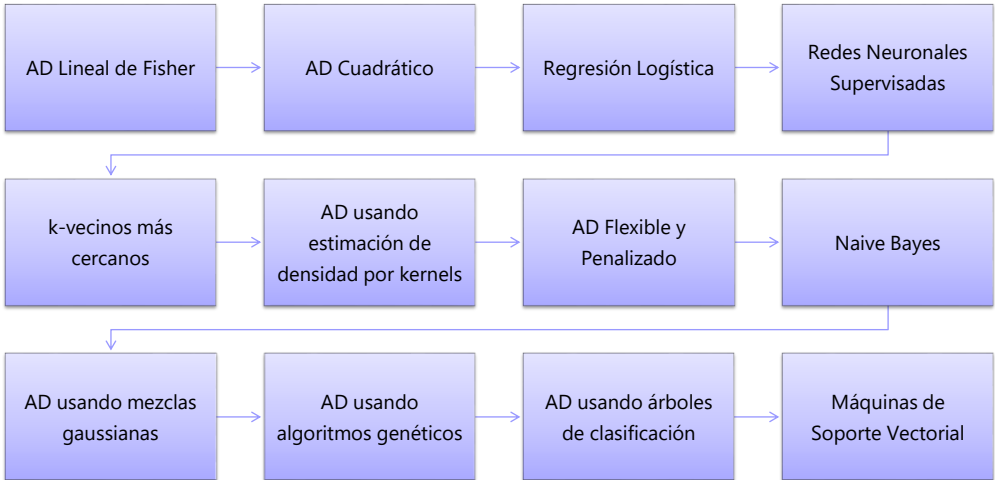
- Se trata de determinar la contribución de cada variable predictora a la clasificación correcta de cada uno de los individuos.

Predictivos

- Se trata de determinar el grupo al que pertenece un individuo para el que se conocen los valores que toman las variables predictoras.



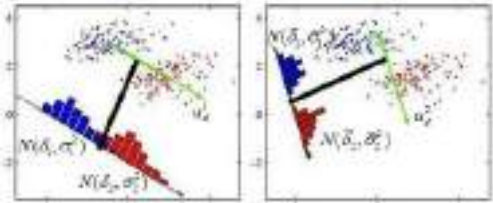
Principales Técnicas Discriminantes



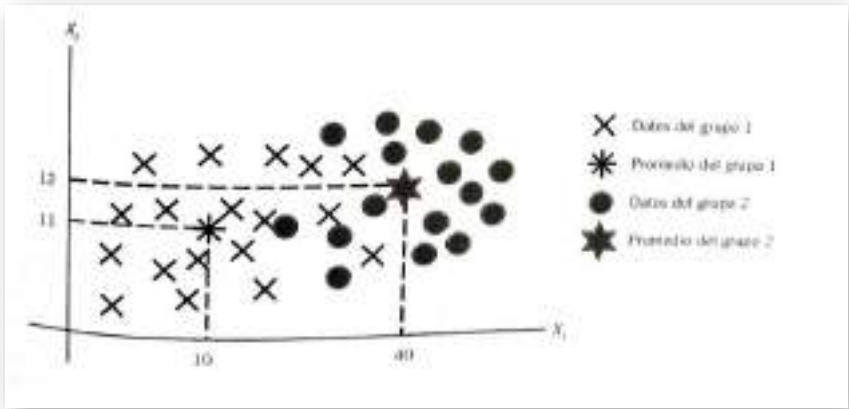
Análisis Discriminante Lineal



Mg. Jesús Salinas Flores

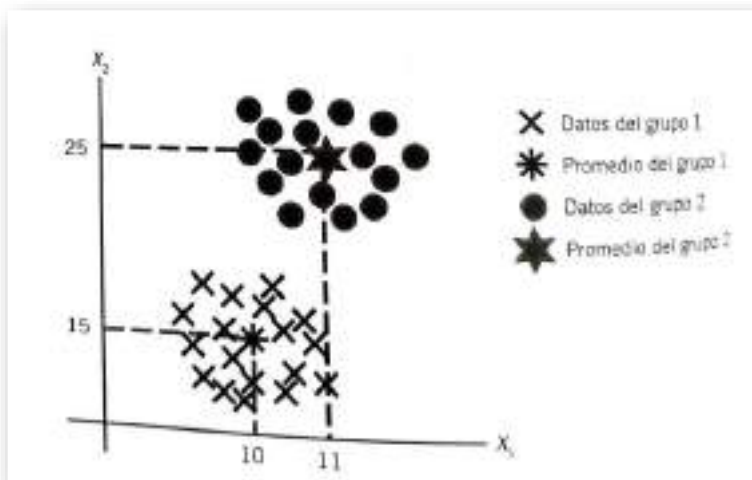


jsalinas@lamolina.edu.pe



Caso 1. Diferencia entre grupos

Fuente: De la Garza, J. & Morales, B. & González, B. (2013). Análisis Estadístico Multivariante. México: Mc Gr



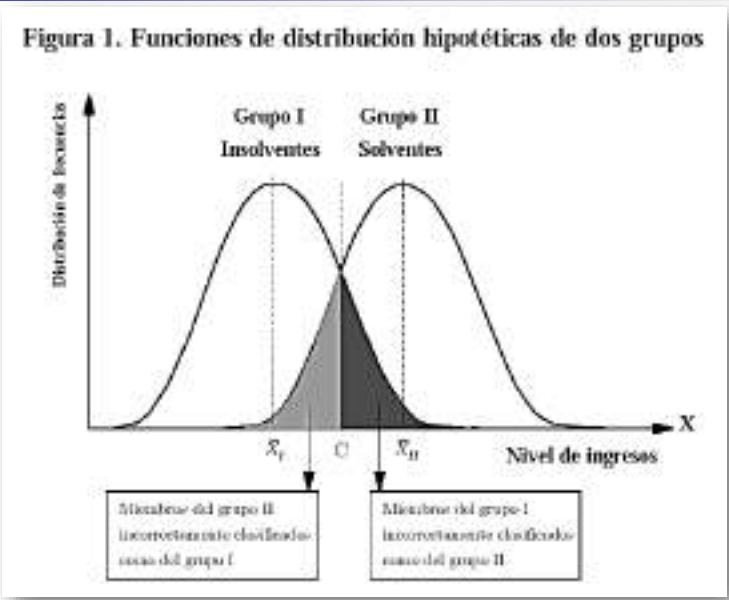
Caso 2. Diferencia entre grupos

Fuente: De la Garza, J. & Morales, B. & González, B. (2013). Análisis Estadístico Multivariante. México: Mc Graw Hill.

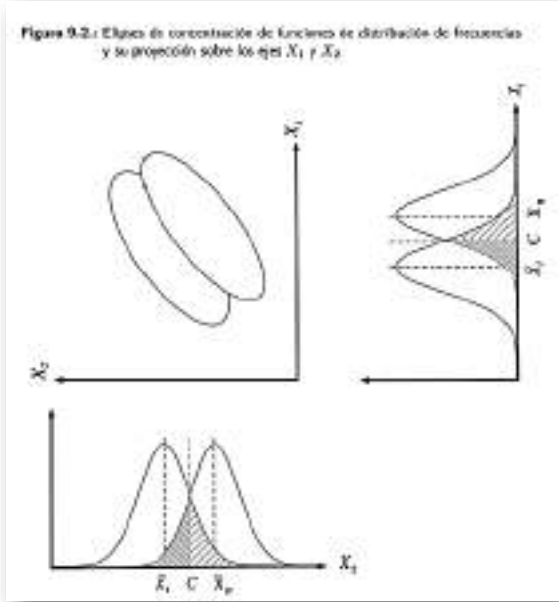


$$\text{Maximización de: } \frac{(\mu_{1,Y} - \mu_{2,Y})}{\sigma_Y^2} = \frac{\text{Variabilidad entre grupos}}{\text{Variabilidad dentro de grupos}}$$

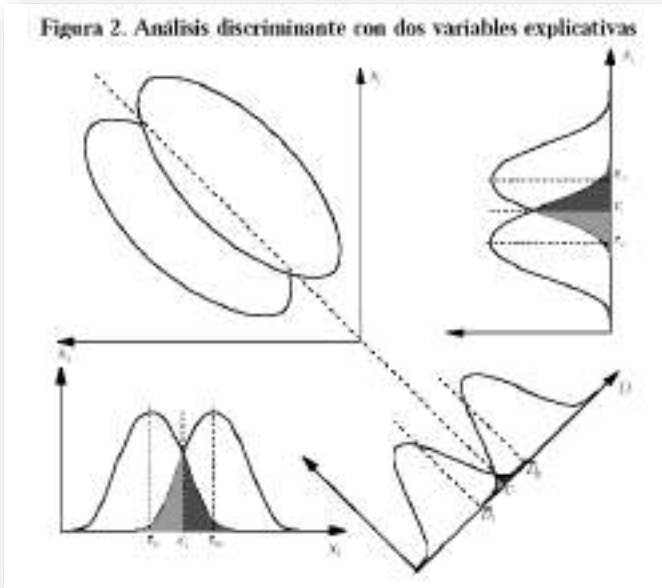
Fuente: De la Garza, J. & Morales, B. & González, B. (2013). Análisis Estadístico Multivariante. México: Mc Graw Hill.



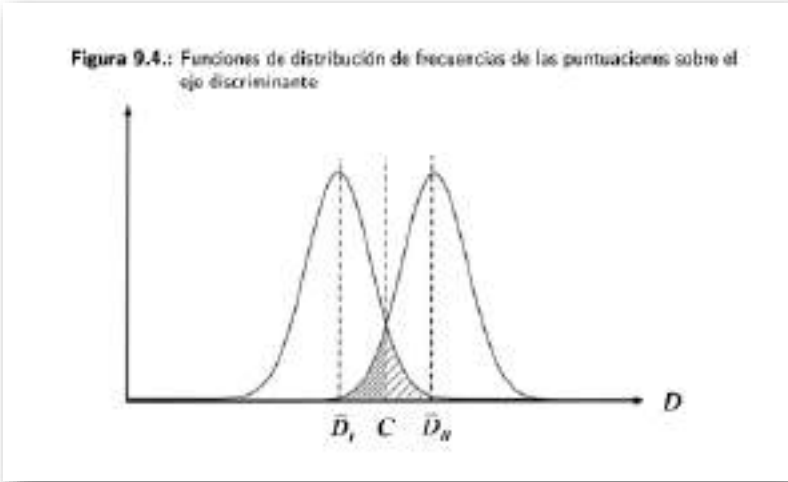
Fuente: Uriel, E. y Aldas, J. (2017). Análisis multivariante aplicado con R. Madrid: Ediciones Paraninfo.



Fuente: Uriel, E. y Aldas, J. (2017). Análisis multivariante aplicado con R. Madrid: Ediciones Paraninfo.

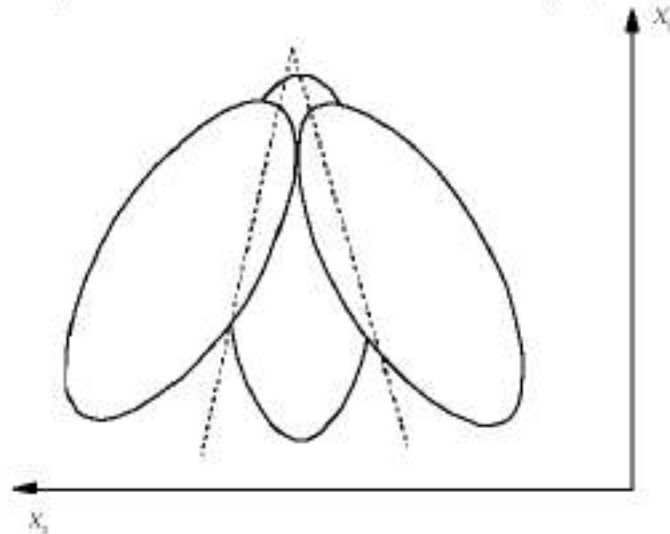


Fuente: Uriel, E. y Aldas, J. (2017). Análisis multivariante aplicado con R. Madrid: Ediciones Paraninfo.



Fuente: Uriel, E. y Aldas, J. (2017). Análisis multivariante aplicado con R. Madrid: Ediciones Paraninfo.

Figura 3. Ilustración del caso de tres grupos

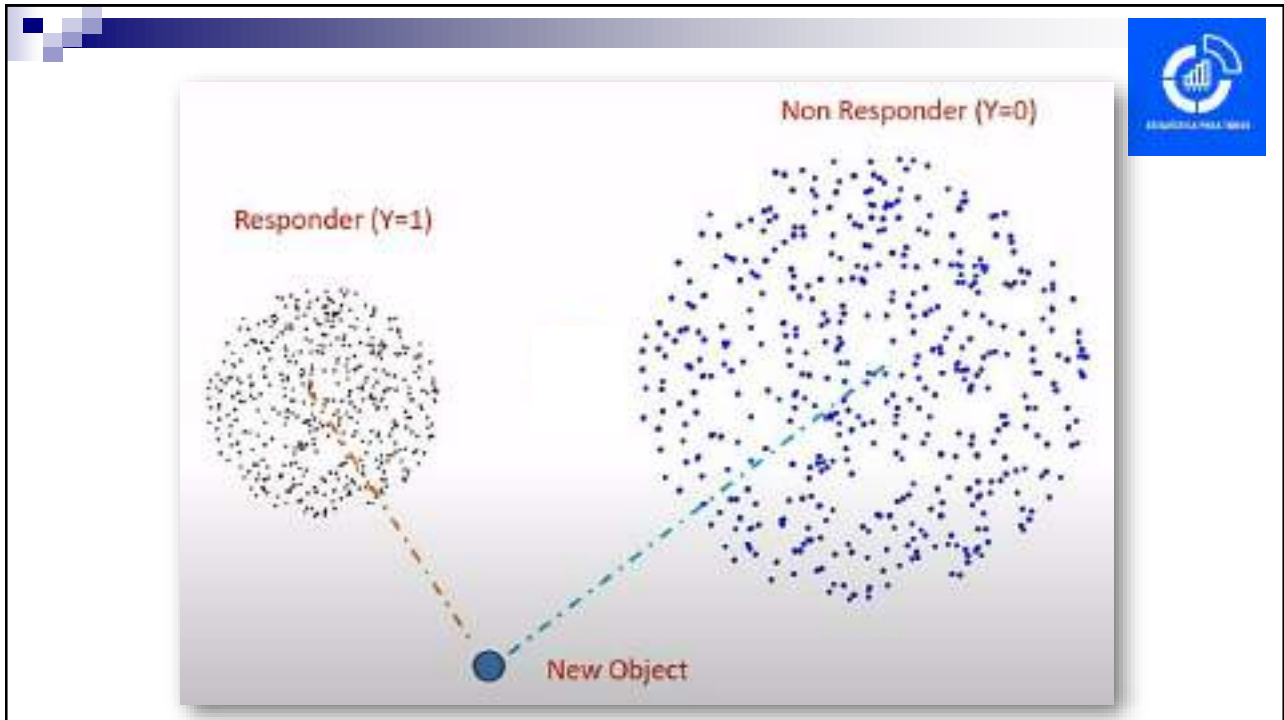


Distancia de Mahalanobis



Mg. Jesús Salinas Flores

jsalinas@lamolina.edu.pe



Distancia Euclidiana y de Mahalanobis

Si hay dos vectores x e y , la distancia euclidiana es $d = \sqrt{\sum (x - y)^2}$

En notación matricial, la distancia euclidiana es $d = \sqrt{\sum (x - y)^T (x - y)}$

Equivalente a $\sqrt{\sum (x - y)^T I (x - y)}$

Se puede ser reemplazado por $\sqrt{\sum (x - y)^T \Sigma^{-1} (x - y)}$

Distancia de Mahalanobis



Transforma las variables en variables no correlacionadas



Hace que sus variancias sean 1



Y luego calcula la distancia euclidiana

$$D_p^2 = (X - \mu)^T \Sigma^{-1} (X - \mu)$$

Criterio de la Distancia de Mahalanobis



Dicha distancia toma en cuenta la correlación entre las variables al tomar en cuenta la matriz de covariancias intra-grupos.

La aplicación de este criterio consiste en asignar cada individuo a la clase para el que la distancia de Mahalanobis sea menor.

$$D_p^2 = (X - \mu)^T \Sigma^{-1} (X - \mu)$$

Esquema metodológico del ADLineal



Ejemplo de Aplicación Suscripción a revista de cine

La compañía de cable edita y promociona una revista de cine (de edición mensual) a un grupo (442) de sus suscriptores durante 6 meses.

Al cabo de dicho periodo le ofrece la posibilidad de suscribirse a dicha revista.

De los 442 clientes a los que se ofreció la promoción, se suscribieron a la revista 329 y no se suscribieron 113



Descripción de las variables



| Variable | Descripción |
|----------------|--|
| Educación | Educación en años |
| Edad | Edad en años |
| Tvdiario | Horas de TV al día durante los fines de semana |
| Organizaciones | Número de clubes al que pertenece |
| Hijos | Número de hijos |
| Suscripción | Decisión de suscribirse |

Consideraciones en la selección de variables (Dallas, 2000)



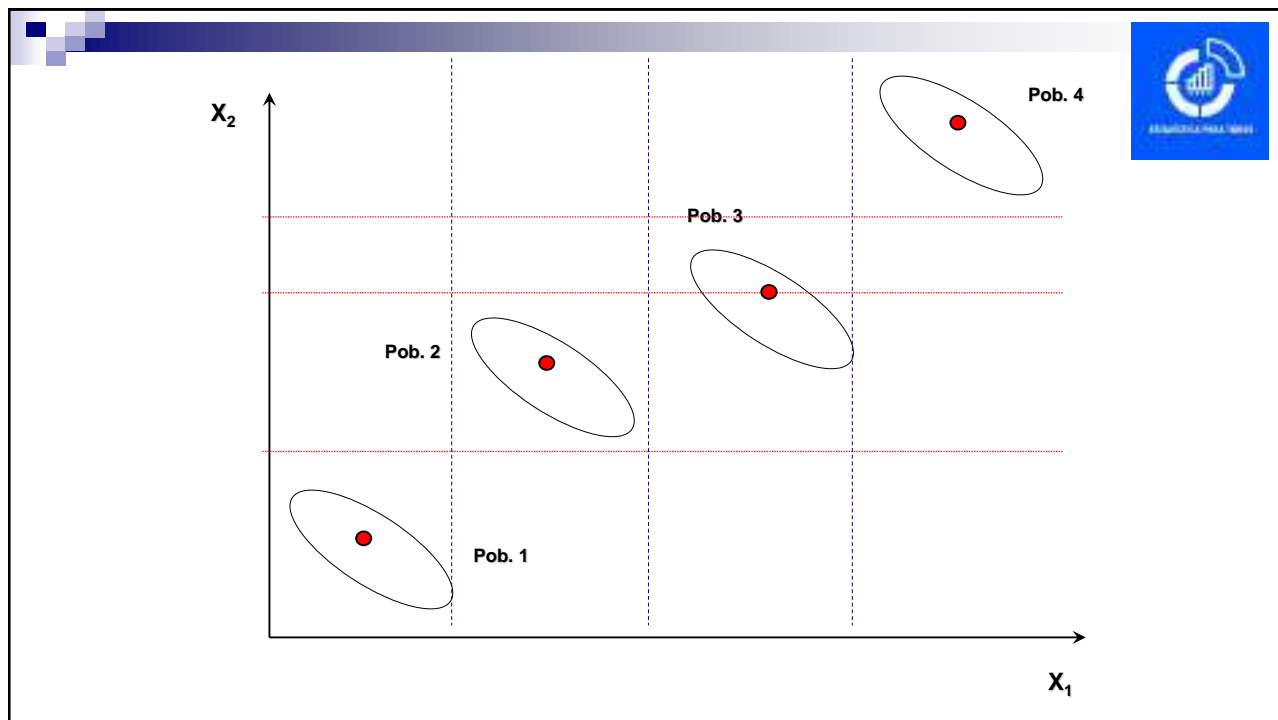
Los métodos de selección se basan en pruebas acerca de las medias de los grupos sobre las variables candidatas y las poblaciones pueden tener media diferentes y, sin embargo, traslaparse.



El que las variables discriminen bien no depende de si las medias son diferentes, sino de si las poblaciones se traslapan.



Sólo porque las poblaciones tengan media diferentes en más de una variable no se concluye la necesidad que se requerirán todas las variables para discriminar adecuadamente



Lambda de Wilks

$$\Lambda = \frac{SC_{Intragrupos}}{SC_{Entregrupos} + SC_{Intragrupos}}$$

La variable que se incluye es la que minimiza la lambda, sin tener en cuenta el efecto de las variables que se hayan podido entrar en el modelo



Estimación de la función discriminante

Obtención de la ecuación asociada a cada función discriminante.



Si la variable a explicar es de m grupos, el AD calcula $m-1$ funciones discriminante.



$$Z_i = b_0 + b_1X_{1i} + \dots + b_pX_{pi}$$



Coeficientes discriminantes b_p

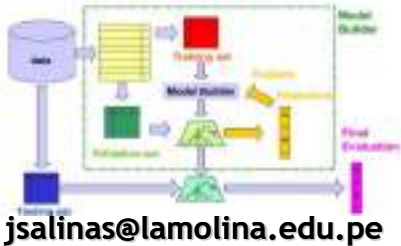
Los coeficientes se estiman teniendo en cuenta que las puntuaciones discriminantes de los individuos de un grupo sean lo más distintas posible de las puntuaciones discriminantes de los individuos de otro grupo.

Esto ocurre cuando el ratio SC_{inter} / SC_{intra} es máximo

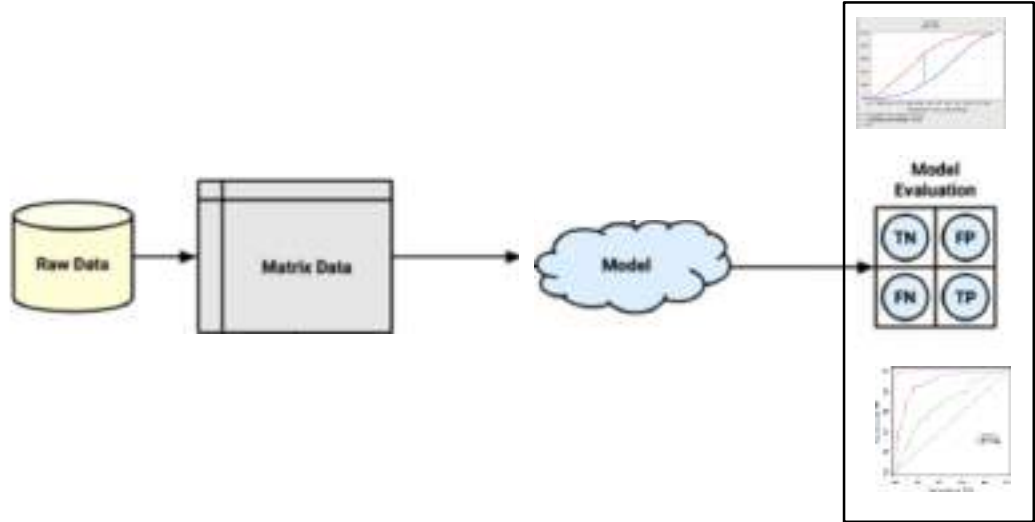
Métodos de Validación



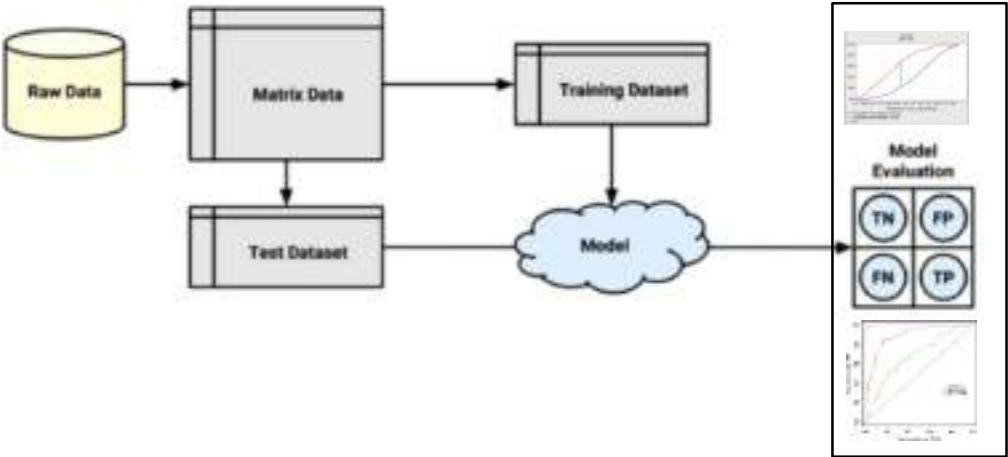
Mg. Jesús Salinas Flores



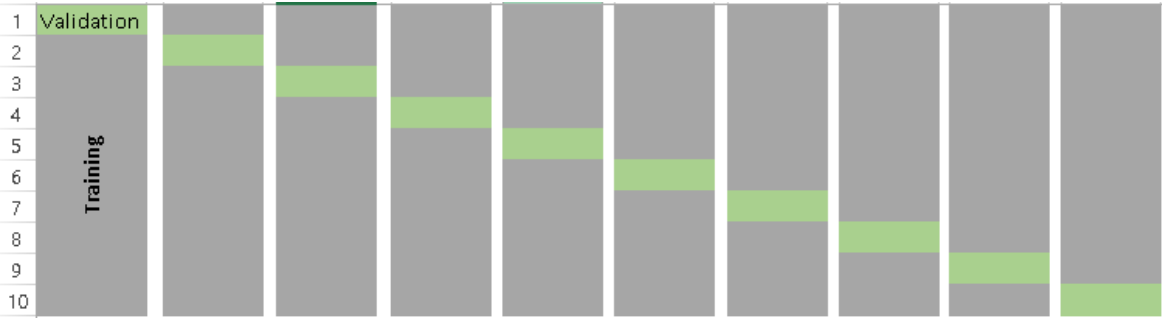
Validación por Resustitución



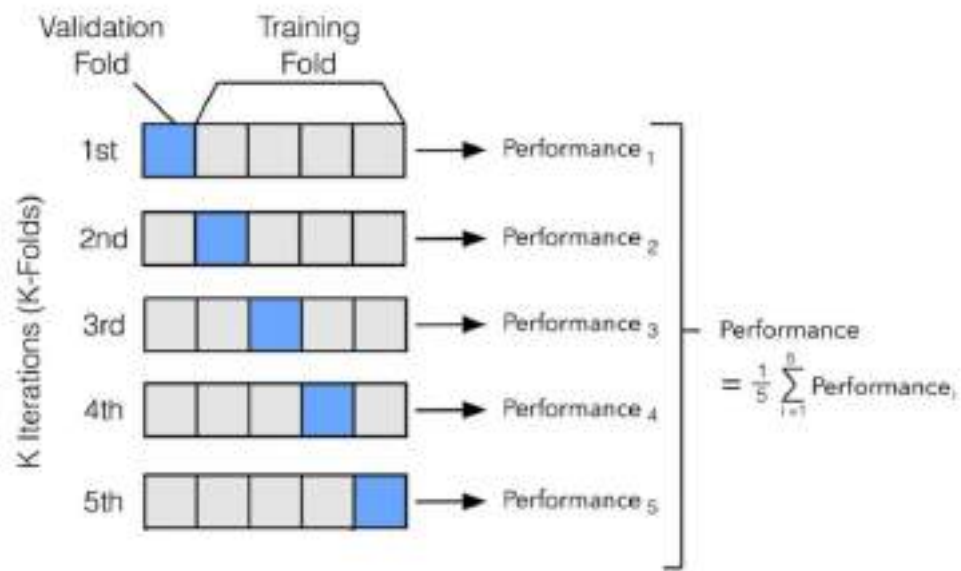
Validación por Conjunto de Prueba



Validación Cruzada



Fuente: https://www.analyticsvidhya.com/blog/2018/05/improve-model-performance-cross-validation-in-python-r/?utm_source=facebook.com&utm_medium=social



Indicadores para evaluación de modelos de clasificación



Mg. Jesús Salinas Flores



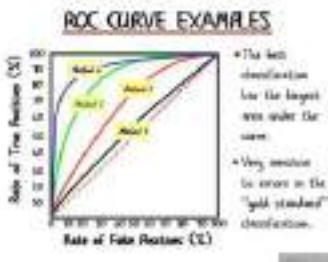
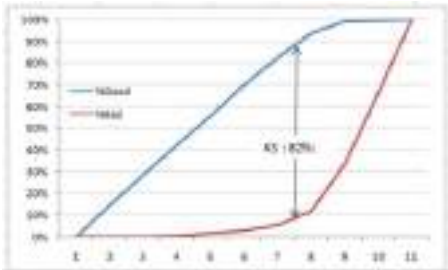
jsalinas@lamolina.edu.pe



Indicadores para evaluar modelos de clasificación



$$K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$



Matriz de Confusión



Mg. Jesús Salinas Flores



Tipos de error al contrastar una hipótesis

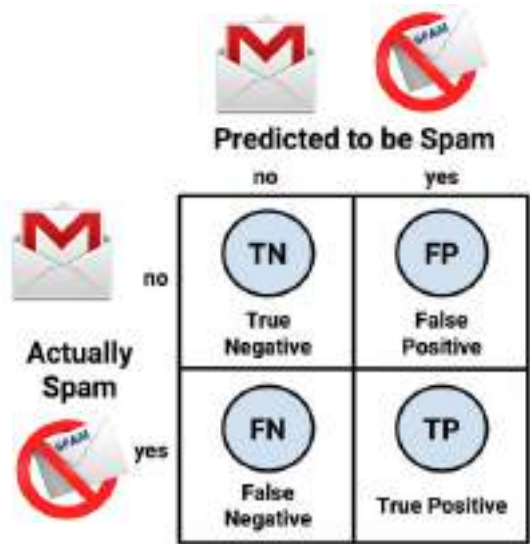
| | | Decisión | |
|----------|------------------------|--|---|
| | | Acepto H_0 (-) | Rechazo H_0 (+) |
| Realidad | H_0 verdadera (-) | Decisión correcta Probabilidad = $1 - \alpha$ | Error de tipo I Probabilidad = α |
| | H_0 falsa (+) | Error de tipo II Probabilidad = β | Decisión correcta Probabilidad = $1 - \beta$ |

Tipos de Errores en un diagnóstico

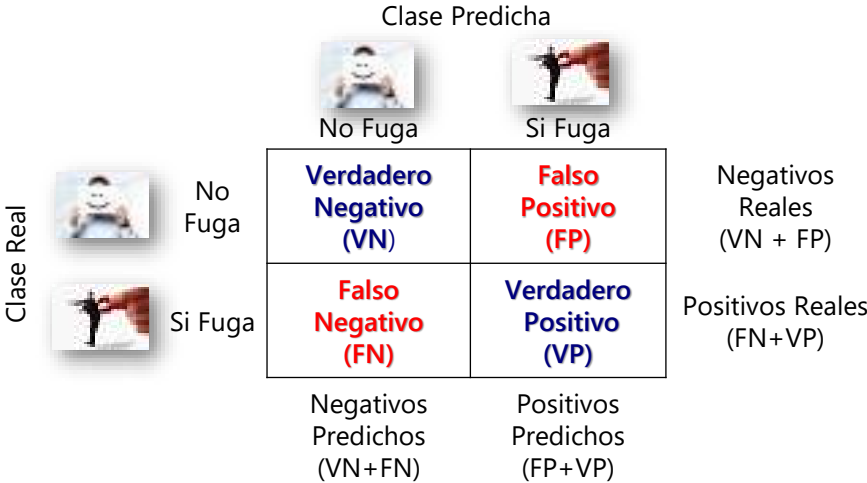




Matriz de Confusión



Matriz de Confusión





Matriz de Confusión

| ID | CHURN | PROBA | CLASE.PRED |
|----|--------|-------|------------|
| 1 | Fuga | 0.658 | Fuga |
| 2 | Actual | 0.089 | Actual |
| 3 | Actual | 0.107 | Actual |
| 4 | Fuga | 0.658 | Fuga |
| 5 | Actual | 0.781 | Fuga |
| 6 | Fuga | 0.720 | Fuga |
| 7 | Actual | 0.636 | Fuga |
| 8 | Actual | 0.834 | Fuga |
| 9 | Actual | 0.829 | Fuga |
| 10 | Fuga | 0.251 | Actual |

| | | Clase Predicha | |
|------------|-----------|----------------|----------|
| | | Actual (-) | Fuga (+) |
| Clase Real | Actual(-) | | |
| | Fuga (+) | | |



Matriz de Confusión

| ID | CHURN | PROBA | CLASE.PRED |
|----|--------|-------|------------|
| 1 | Fuga | 0.658 | Fuga |
| 2 | Actual | 0.089 | Actual |
| 3 | Actual | 0.107 | Actual |
| 4 | Fuga | 0.658 | Fuga |
| 5 | Actual | 0.781 | Fuga |
| 6 | Fuga | 0.720 | Fuga |
| 7 | Actual | 0.636 | Fuga |
| 8 | Actual | 0.834 | Fuga |
| 9 | Actual | 0.829 | Fuga |
| 10 | Fuga | 0.251 | Actual |

| | | Clase Predicha | |
|------------|-----------|----------------|----------|
| | | Actual (-) | Fuga (+) |
| Clase Real | Actual(-) | | |
| | Fuga (+) | | |



Matriz de Confusión

| | | Clase Predicha | | |
|------------|---|--|--|----------------------------------|
| | | 0 | 1 | |
| Clase Real | 0 | Verdadero Negativo (VN) | Falso Positivo (FP) | Negativos Reales (VN + FP) |
| | 1 | Falso Negativo (FN) | Verdadero Positivo (VP) | Positivos Reales (FN+VP) |
| | | Negativos Predichos (VN+FN) | Positivos Predichos (FP+VP) | |



Matriz de Confusión

- $\text{Accuracy} = (\text{VP} + \text{VN}) / \text{Total de casos}$
- $\text{Tasa de error} = (\text{FP} + \text{FN}) / \text{Total de casos} = 1 - \text{Accuracy}$
- $\text{Sensibilidad} = \text{Verdadero Positivo} / \text{Positivos Reales} = \text{VP} / (\text{FN} + \text{VP})$
- $\text{Especificidad} = \text{Verdadero Negativo} / \text{Negativos Reales} = \text{VN} / (\text{VN} + \text{FP})$
- $\text{Precisión} = \text{Valor Predictivo Positivo (VP+)} = \text{Verdadero Positivo} / \text{Positivos Predichos} = \text{VP} / (\text{FP} + \text{VP})$
- $\text{Valor Predictivo Negativo (VP-)} = \text{Verdadero Negativo} / \text{Negativos Predichos} = \text{VN} / (\text{VN} + \text{FN})$



Sensibilidad y VP+

Sensibilidad = Verdaderos Positivos/Positivos reales
Valor predictivo positivo (VP+) = Verdaderos positivos/positivos predichos

| | | Clase Predicha | | |
|------------|---|-----------------------------|-------------------------|-----------------------------|
| | | 0 | 1 | |
| Clase Real | 0 | | | Positivos Reales (FN+VP) |
| | 1 | | Verdadero Positivo (VP) | |
| | | Positivos Predichos (FP+VP) | | |

Detección fraude: Sensibilidad
Marketing directo: VP+



Especificidad y VP-

Especificidad = Verdadero negativo/Negativo Real
Valor predictivo negativo (VP-) = Verdadero negativo/Negativo Predicho

| | | Clase Predicha | | |
|------------|---|-----------------------------|---|-----------------------------|
| | | 0 | 1 | |
| Clase Real | 0 | Verdadero Negativo (VN) | | Negativos Reales (VN+FP) |
| | 1 | | | |
| | | Negativos Predichos (VN+FN) | | |

Recuperación de información: Especificidad
Riesgo: VP-

Sensibilidad, Especificidad y Tasa de Detección



- ❑ la **sensibilidad** es el porcentaje de los verdaderos 1 que se predijeron correctamente. Muestra qué porcentaje de 1's fueron cubiertos por el modelo. La sensibilidad es más importante cuando clasificar correctamente los 1 es más importante que clasificar los 0's.
- ❑ La **especificidad** es la proporción de los verdaderos 0 que se predijeron correctamente. La especificidad es más importante cuando clasificar correctamente los 0 es más importante que clasificar los 1's.
- ❑ La maximización de la especificidad es más relevante en casos como la detección de spam, en la que, estrictamente, no queremos que los mensajes genuinos (0) terminen en spam (1).
- ❑ La **tasa de detección** es la proporción de toda la muestra en la que los eventos 1 se detectaron correctamente

Precisión, Recall y F1 Score



- ❑ La **precisión** lo que trata es determinar qué porcentaje de las predicciones positivas del modelo (1) son exactas.
- ❑ Si se tiene un modelo con alta precisión, también se quiere saber qué porcentaje de todos los 1's fueron cubiertos. Esto se puede capturar usando la Sensibilidad. Pero en este contexto, se conoce como **Recall**.
- ❑ Un puntaje de alta precisión da más confianza a la capacidad del modelo para clasificar los 1's. La combinación de esto con Recall da una idea de cuántos de los 1 totales fue capaz de cubrir.
- ❑ Un buen modelo debe tener una buena precisión, así como un alto Recall. Así que idealmente, se quiere tener una medida que combina estos dos aspectos en una sola métrica - la **Puntuación F1**.
- ❑ Puntaje F1 = $(2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$

Estadístico Kappa



Mg. Jesús Salinas Flores

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

jsalinas@lamolina.edu.pe



| | | Doctor 2 | | |
|----------|-----------------|-----------------|-----------------|----|
| | | No tiene cáncer | Si tiene Cáncer | |
| Doctor 1 | No tiene cáncer | 8 | 0 | 10 |
| | Si tiene cáncer | 0 | 2 | |

| | | Doctor 2 | | |
|----------|-----------------|-----------------|-----------------|----|
| | | No tiene cáncer | Si tiene Cáncer | |
| Doctor 1 | No tiene cáncer | 6 | 2 | 8 |
| | Si tiene cáncer | 1 | 1 | 2 |
| | | | | 10 |



Estadístico Kappa

- El estadístico **Kappa** es similar a la accuracy, pero tiene en cuenta la posibilidad de una predicción correcta por casualidad.
- **Kappa = (Observed Accuracy - Expected Accuracy) / (1 - Expected Accuracy)**
- ❖ Los valores de Kappa oscilan entre 0 y un máximo de 1, lo que indica una concordancia perfecta entre las predicciones del modelo y los valores reales. Los valores inferiores a uno indican una concordancia imperfecta.

Interpretación de los valores de kappa

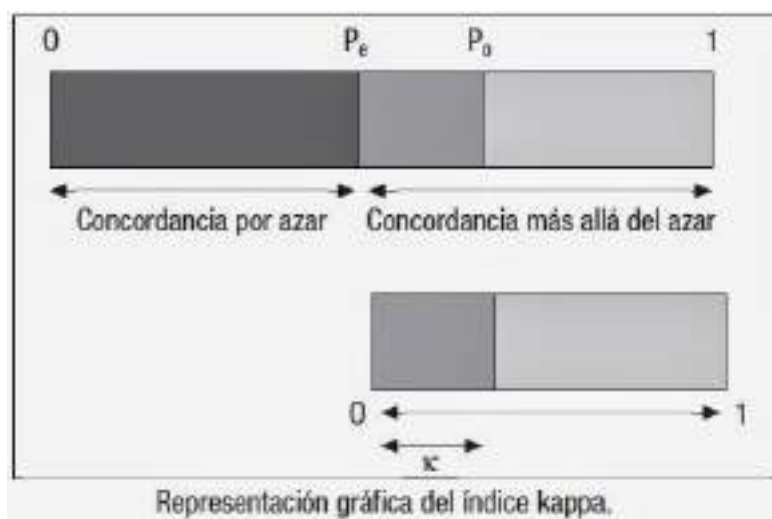
| Valor de kappa | Grado de concordancia |
|----------------|-----------------------|
| 0,81-1,00 | Excelente |
| 0,61-0,80 | Buena |
| 0,41-0,60 | Moderada |
| 0,21-0,40 | Ligera |
| < 0,20 | Mala |

Estadístico Kappa



$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Definición de Kappa según Cohen





Estadístico Kappa (Caso 1)

| Detección de Correo Spam | | Predicho | | Total |
|--------------------------|---------|----------|------|-------|
| | | No Spam | Spam | |
| Observado | No Spam | 850 | 50 | 900 |
| | Spam | 30 | 70 | 100 |
| Total | | 880 | 120 | 1000 |

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Definición de Kappa según Cohen

$$P_o = \frac{(850+70)}{1000} = 0,92$$

$$\kappa = \frac{(0,92 - 0,804)}{(1 - 0,804)} = 0,5918$$

$$P_e = \underbrace{\frac{900}{1000} * \frac{880}{1000}}_{\text{No Spam}} + \underbrace{\frac{100}{1000} * \frac{120}{1000}}_{\text{Spam}} = 0,804$$



Estadístico Kappa (Caso 2)

| Detección de Correo Spam | | Predicho | | Total |
|--------------------------|---------|----------|------|-------|
| | | No Spam | Spam | |
| Observado | No Spam | 900 | 0 | 900 |
| | Spam | 0 | 100 | 100 |
| Total | | 900 | 100 | 1000 |

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Definición de Kappa según Cohen

$$P_o = \frac{(900+100)}{1000} = 1$$

$$\kappa = \frac{(1 - 0,82)}{(1 - 0,82)} = 1$$

$$P_e = \underbrace{\frac{900}{1000} * \frac{900}{1000}}_{\text{No Spam}} + \underbrace{\frac{100}{1000} * \frac{100}{1000}}_{\text{Spam}} = 0,82$$

Estadístico Kappa (Caso 3)



| Detección de Correo Spam | | Predicho | | Total |
|--------------------------|---------|----------|------|-------|
| | | No Spam | Spam | |
| Observado | No Spam | 400 | 500 | 900 |
| | Spam | 60 | 40 | 100 |
| Total | | 460 | 540 | 1000 |

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Definición de Kappa según Cohen

$$P_o = \frac{(400+40)}{1000} = 0,44$$

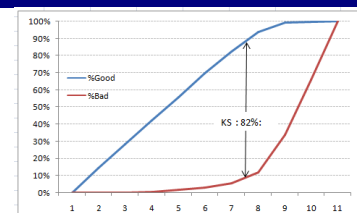
$$\kappa = \frac{(0,44 - 0,468)}{(1 - 0,468)} = -0,0526$$

$$P_e = \underbrace{\frac{900}{1000} * \frac{460}{1000}}_{\text{No Spam}} + \underbrace{\frac{100}{1000} * \frac{540}{1000}}_{\text{Spam}} = 0,468$$

Estadístico K-S



Mg. Jesús Salinas Flores



jsalinas@lamolina.edu.pe

Estadístico Kolgomorov-Smirnov (KS)



- El estadístico KS ayuda a entender, qué porción de la población debe ser dirigida para obtener la mayor tasa de respuesta (1). El KS se usa para tomar decisiones como: ¿Cuántos clientes deberíamos apuntar para dirigir una campaña de marketing? o ¿A cuántos clientes debemos pagar para mostrar anuncios, etc.
- El gráfico y el estadístico KS se utiliza ampliamente en los escenarios de credit scoring para seleccionar el tamaño óptimo de la población de los usuarios objetivo para las campañas de marketing.

Estadístico Kolgomorov-Smirnov (KS)



- **Paso 1:** Una vez obtenidas las puntuaciones de probabilidad de predicción, las observaciones se ordenan por orden decreciente de las puntuaciones de probabilidad. De esta manera, puede esperar que las filas en la parte superior se clasifiquen como 1 mientras que las filas en la parte inferior son 0's.
- **Paso 2:** Todas las observaciones se dividen en 10 partes de igual tamaño (deciles).
- **Paso 3:** Luego, el estadístico de KS es la diferencia máxima entre el porcentaje acumulado de respuesta o 1 (tasa positiva verdadera acumulada) y el porcentaje acumulativo de no respuesta o 0 (tasa positiva falsa acumulada).



Estadístico Kolgomorov-Smirnov (KS)

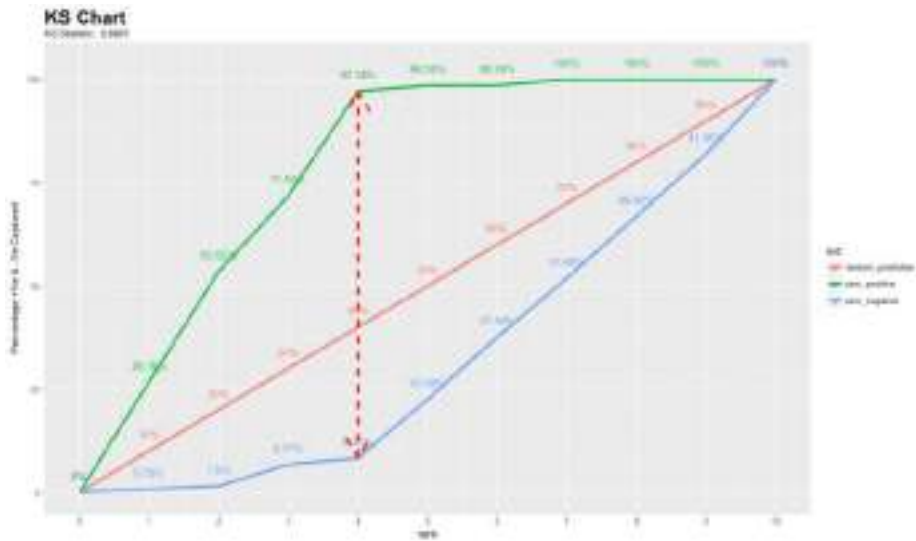
- La tabla K-S o Kolmogorov-Smirnov mide el desempeño de los modelos de clasificación.
- K-S es una medida del grado de separación entre las distribuciones positiva y negativa.
- El K-S es 100, si las puntuaciones dividen la población en dos grupos separados en los que un grupo contiene todos los positivos y el otro todos los negativos.
- Por otro lado, si el modelo no puede diferenciar entre positivos y negativos, entonces es como si el modelo seleccionara casos aleatoriamente de la población.
- El K-S sería 0. En la mayoría de los modelos de clasificación, el K-S caerá entre 0 y 100, y que cuanto mayor sea el valor, mejor será el modelo para separar los casos positivos de negativos.



Estadístico Kolgomorov-Smirnov (KS)

| ks.stat(y_pred, y_oct, returnKStab=c=T) | | | | | | | | | | |
|---|------|-----------|----------------|------------|-------------------------------|-----------------|---------------------|----------------------|--------------------------|------------|
| | rank | total_pos | non_responders | responders | expected_responders_by_random | perc_responders | perc_non_responders | calc_perc_responders | calc_perc_non_responders | difference |
| 1 | 1 | 1 | 20 | 0 | 20 | 7.041376 | 0.24681358 | 0.00000000 | 0.24681358 | 0.00000000 |
| 2 | 2 | 2 | 20 | 0 | 20 | 7.041376 | 0.24681358 | 0.00000000 | 0.49362716 | 0.00000000 |
| 3 | 3 | 3 | 20 | 1 | 19 | 7.041376 | 0.25459750 | 0.00138881 | 0.50598591 | 0.00138881 |
| 4 | 4 | 4 | 20 | 3 | 17 | 7.041376 | 0.14834815 | 0.00504050 | 0.37154132 | 0.37154132 |
| 5 | 5 | 5 | 20 | 20 | 0 | 7.041376 | 0.00000000 | 0.16248162 | 0.57694332 | 0.57694332 |
| 6 | 6 | 6 | 20 | 18 | 2 | 7.041376 | 0.82469132 | 0.14634463 | 0.9812346 | 0.30211321 |
| 7 | 7 | 7 | 20 | 15 | 5 | 7.041376 | 0.89070890 | 0.12239120 | 0.9636030 | 0.38480304 |
| 8 | 8 | 8 | 20 | 10 | 10 | 7.041376 | 0.81234568 | 0.15447154 | 0.9753000 | 0.05836030 |
| 9 | 9 | 9 | 20 | 20 | 0 | 7.041376 | 0.00000000 | 0.16248162 | 0.9753000 | 0.30211321 |
| 20 | 10 | 24 | 22 | 2 | 9.525412 | 0.82469132 | 0.17850170 | 1.0000000 | 1.00000000 | 0.00000000 |

Gráfico Kolgomorov-Smirnov (KS)

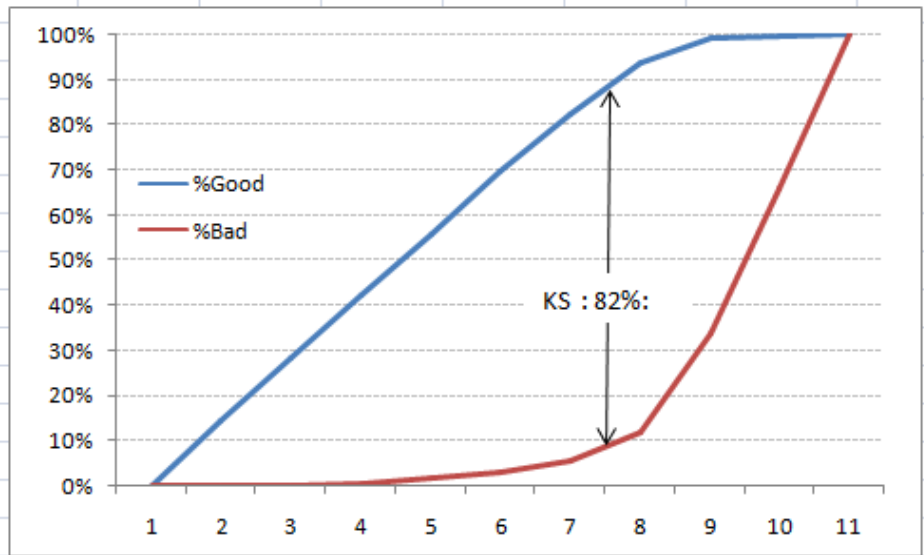


Estadístico Kolgomorov-Smirnov (KS)

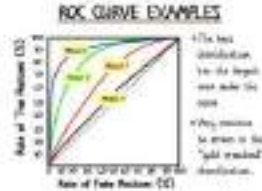


| Lift/Gain Column | | | | | | Cummulative | | K-S |
|------------------|------|------|-----------|---------|---------|-------------|------------|---------|
| Row La | 0 | 1 | Grand Tot | %Rights | %Wrongs | Cum %Rig | Cum %Wrong | |
| 1 | 543 | 543 | | 14% | 0% | 14% | 0% | 14% |
| 2 | 2 | 542 | 544 | 14% | 0% | 28% | 0% | 28% |
| 3 | 7 | 537 | 544 | 14% | 0% | 42% | 1% | 42% |
| 4 | 15 | 529 | 544 | 14% | 1% | 56% | 2% | 54% |
| 5 | 20 | 524 | 544 | 14% | 1% | 69% | 3% | 67% |
| 6 | 42 | 502 | 544 | 13% | 3% | 83% | 5% | 77% |
| 7 | 104 | 440 | 544 | 11% | 7% | 94% | 12% | 82% K-S |
| 8 | 345 | 199 | 544 | 5% | 22% | 99% | 34% | 65% |
| 9 | 515 | 29 | 544 | 1% | 32% | 100% | 66% | 34% |
| 10 | 540 | 5 | 545 | 0% | 34% | 100% | 100% | 0% |
| Grand Tot | 1590 | 3850 | 5440 | | | | | |

Gráfico Kolgomorov-Smirnov (KS)



Curva ROC



Mg. Jesús Salinas Flores

jsalinas@lamolina.edu.pe

Curva ROC



La curva de característica operativa (Receiver Operating Characteristic) permite cuantificar la precisión discriminatoria de un modelo.



Permite evaluar la realización de esquemas de clasificación en los que exista una variable con dos categorías por las que se clasifiquen los individuos.

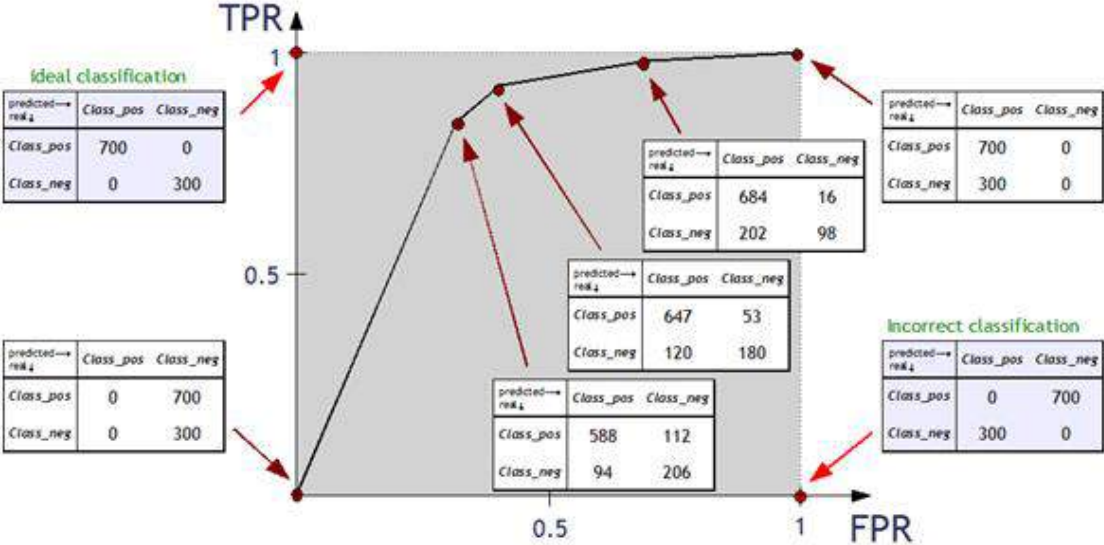
Forma de cálculo de la curva ROC



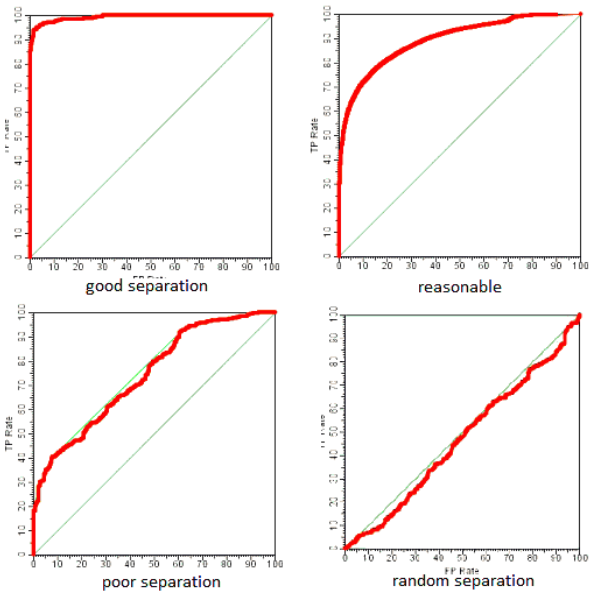
1. Calcular los valores de la función de decisión.
2. Se prueba el clasificador para diferentes puntos de corte (umbrales). El alfa es el umbral de la probabilidad estimada, por encima de la cual se asigna una observación a una categoría (clase positiva) y debajo a la otra categoría (clase negativa).
3. Para cada clasificación con un valor del umbral se obtiene un par (TPR, FPR), que corresponde a un punto en la curva ROC.
4. Para cada clasificación con un valor del umbral también se obtiene la matriz de confusión correspondiente.



Curva ROC

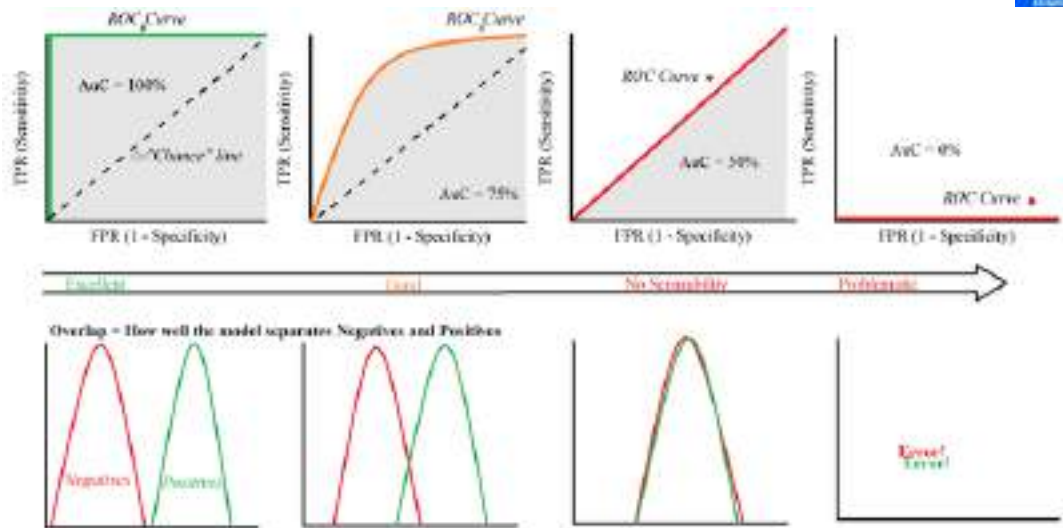


Curva ROC





Curva ROC



Fuente: <https://www.datasciencecentral.com/profiles/blogs/roc-curve-explained-in-one-picture>

Coeficiente de Gini



Mg. Jesús Salinas Flores

jsalinas@lamolina.edu.pe

Coeficiente de Gini



- ❖ El Coeficiente de Gini es un indicador de lo bien que el modelo supera las predicciones aleatorias.
- ❖ Se puede calcular a partir del área bajo la curva ROC usando la siguiente fórmula:
- ❖ Coeficiente de Gini = $(2 * AUROC) - 1$

Log-Loss



$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

Mg. Jesús Salinas Flores

jsalinas@lamolina.edu.pe



Log Loss

- ❖ Log Loss cuantifica la precisión de un clasificador al penalizar las clasificaciones falsas.
- ❖ Minimizar el Log Loss es básicamente equivalente a maximizar la exactitud del clasificador, pero con un pequeño detalle.
- ❖ Para calcular el Log Loss, el clasificador debe asignar una probabilidad a cada clase.



Log Loss

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

Donde:

- ❖ N es el número de muestras o instancias, M es el número de etiquetas posibles, y_{ij} es un indicador binario de si la etiqueta j es la clasificación correcta para la instancia i, y p_{ij} es la probabilidad del modelo de asignar la etiqueta j a la instancia i.
- ❖ Un clasificador perfecto tendría una pérdida de registro de precisamente cero. Los clasificadores menos ideales tienen valores progresivamente más grandes de Log Loss.

Log Loss

- ❖ Si solo hay dos clases, la expresión anterior simplifica a:

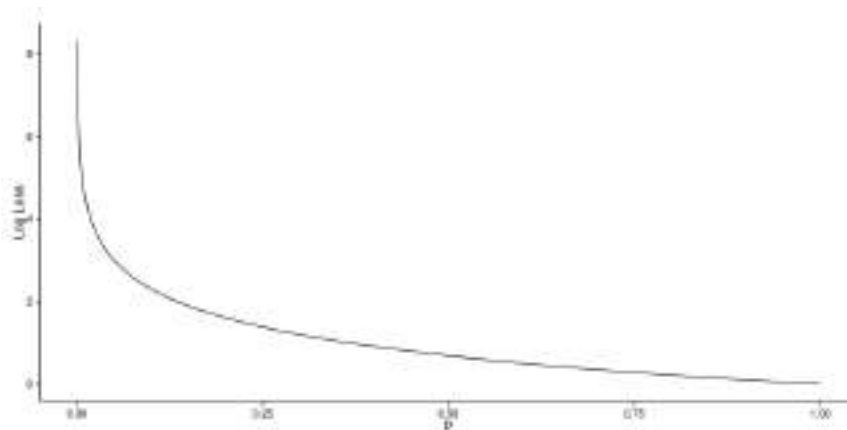
$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

- ❖ Tenga en cuenta que para cada instancia solo el término para la clase correcta realmente contribuye a la suma.



Log Loss

A medida que la probabilidad predicha de la clase verdadera se acerca a cero, la pérdida aumenta exponencialmente:



Log Loss



- ❖ Log Loss penaliza fuertemente a los clasificadores que confían en una clasificación incorrecta.
- ❖ Por ejemplo, si para una observación particular, el clasificador asigna una probabilidad muy pequeña a la clase correcta, entonces la contribución correspondiente al Log Loss será muy significativa.
- ❖ La conclusión es que es mejor estar algo equivocado que enfáticamente equivocado. Por supuesto, siempre es mejor estar completamente en lo correcto