

The background is a blurred image of a financial market display. It features various stock indices and their values, such as 'OMX COPENHAGEN 25 INDEX' at 1172.94, 'OMXRG1' at 984.13, and 'OMX ICELAND 8' at 6230.9. There are also line charts showing market trends. The overall color scheme is dominated by blue and red, typical of financial data visualizations.

Análisis Discriminante

Dr. Misael Erikson Maguiña Palma



Análisis discriminante

Se utiliza para caracterizar grupos de consumidores cuando se conoce los grupos de pertenencia, por eso es una técnica de clasificación.

- Análisis discriminante de **dos grupos** (simple),
- Análisis discriminante múltiple (**mas de tres categorías**).



Análisis Discrimínate

Objetivos:

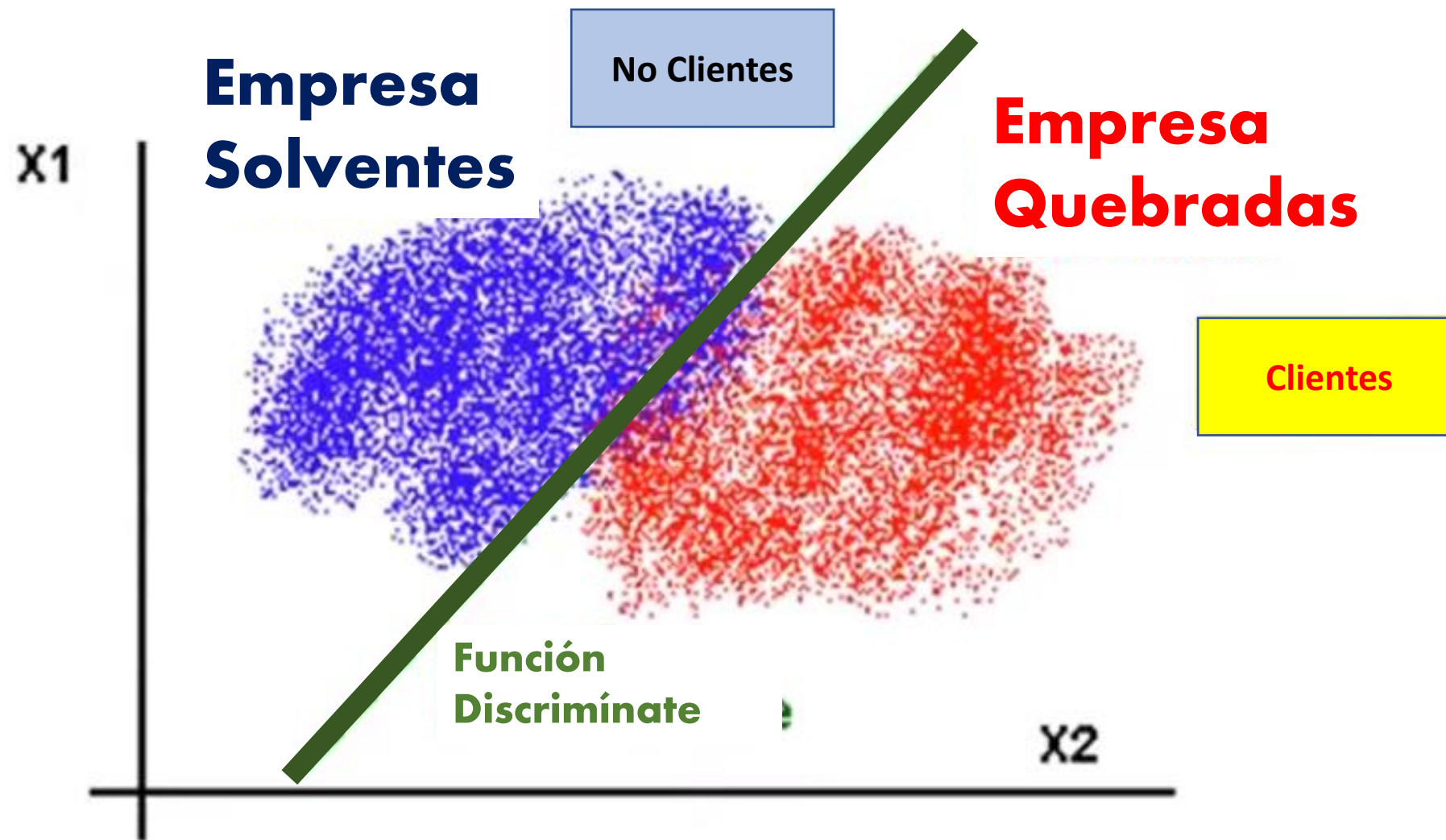
- Separación o discriminación entre grupos (objetivo de carácter descriptivo) tratando de encontrar las diferencias entre dos o mas grupos atreves de una función discriminante (generalmente lineal) de las variables de identificación
- Predicción o asignación de un objeto o individuo inferencia).



Análisis Discriminante

- Variables dependientes categóricas (puede ser dummy)
- Variables independientes cuantitativas,
- Se cuenta con una tabla de datos de dimensiones “n x p” con “n” el número de variables ubicadas en las columnas,
- Se trata de reducir la dimensión de esta tabla en las columnas, de tal manera que al proyectar la nube de puntos en el espacio, los individuos del mismo grupo se encuentran cerca entre si y los de grupos diferentes estén lo más separado posible.

Que buscamos con el análisis discriminante



Análisis Discrimínate

- La puntuación discriminante para el i-ésimo,

Individuos en la función discriminante “s” será:

$$D_{si} = B_{s1}X_{i1} + \dots + B_{s2}X_{ip} + B_{s0}$$

D= Calificación discriminante,
B=coeficiente o peso discriminante,
X=variable predictiva o independiente.

Con base en estas puntuaciones discriminantes se clasifica el individuo “i” (del cual se sabe de antemano a que grupo pertenece). El porcentaje de correctamente clasificados será un índice de efectividad de las funciones discriminantes.



Supuestos del análisis discriminante

- Las variables dependientes es categórica,
- Las variables independientes son métricas u ordinales,
- Mediciones son independientes,
- Las variables independientes no presentan outliers significativos,
- Las variables independientes son iguales en todos los grupos,
- Las covarianzas de las variables independientes son iguales en dos grupos,
- No existe multicolinealidad.



Elección de las variables

Se debe considerar el subconjunto de éstas tal que al representar el subconjunto de éstas en un subespacio generado por los valores de la variable, se evidencien grupos muy separados en los cuales sus elementos están poco dispersos entre si muy cercanos al centro



Lambda de Wilks

- Para un conjunto de “p” variables independientes mide las desviaciones dentro de cada grupo respecto a las desviaciones totales sin distinguir grupos.
- Si su valor es pequeño, la variabilidad total se deberá a las diferencias entre grupos y, por lo tanto las variables seleccionas discriminan adecuadamente los grupos.



Método de selección por pasos:

- El criterio de selección es encontrar en cada paso la variable que junto con las previamente seleccionadas haga mínimo el valor de LAMBDA DE WILKS.
- Una variable previamente seleccionada es susceptible de ser eliminada en un paso posterior



F de entrada, f de salida y Tolerancia

- El criterio para la selección de una variable será su aporte marginal en la disminución del valor de LW.
- Esta disminución será significativa si el estadístico F presenta un valor mayor a determinado valor crítico. Caso en el cual se selecciona la variable evaluada.
- Una variable preseleccionada será eliminada si el valor del estadístico F presenta un valor menor a determinado valor crítico. Si el valor de F es pequeño, la disminución en LW no contribuye, en el cual será adecuado la eliminación.



La Clasificación

- Se realiza a partir de las funciones discriminantes tipificadas (coeficientes estandarizados de las funciones discriminantes canónicas).

Matriz de confusión

Para poder utilizar este método de evaluación de un modelo de clasificación necesitamos separar nuestra data de entrenamiento en dos datasets. (Train (80%), Test (20%))

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Precisión

Cuando predice positivos, que porcentaje clasifica correctamente?

$$\text{Precisión} = \frac{VP}{\text{Total clasificados positivos}}$$

Valor de predicción negativo

Cuando predice negativo, que porcentaje clasifica correctamente?

$$VPN = \frac{VN}{\text{Total clasificados negativos}}$$

Exactitud (Accuracy)

En general, que porcentaje de la data clasifica correctamente?

$$\text{Exactitud} = \frac{VP + VN}{\text{Total}}$$

Tasa de error (Misclassification Rate)

En general, que porcentaje de la data clasifica incorrectamente?

$$\text{Tasa de error} = \frac{FP + FN}{\text{Total}}$$

Sensibilidad, exhaustividad, Tasa de verdaderos positivos

Traducción al inglés,

Cuando la clase es positiva, que porcentaje logra clasificar?

$$\text{Sensibilidad} = \frac{VP}{\text{Total Positivos}}$$

Especificidad, tasa de verdaderos negativos

Cuando la clase es negativa, que porcentaje logra clasificar?

$$\text{Especificidad} = \frac{VN}{\text{Total Negativos}}$$

