https://media-exp3.licdn.com/dms/image/

RECONOCIMIENTO DE PATRONES EN IMÁGENES TICS 585

FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

SEGUNDO SEMESTRE 2021

PROFESOR: MIGUEL CARRASCO

CLASIFICADORES SUPERVISADOS

- Métodos
  - Min-Max
  - Media cero, desviación uno

OBJETIVO:
Transformar el rango de los datos originales a un nuevo intervalo.

- Técnicas de normalización

Min-Max

Media desviación

- El método de normalización *min-max* transforma el rango de valores originales a un rango fijo contenido entre cero y uno.

- Min-Max

$0 \leq \mathbf{v}_j \leq 1$

$$\mathbf{v}_i = f_N\left(\mathbf{u}_i\right) = \frac{\left(\mathbf{u}_i - \min(\mathbf{u}_i)\right)}{\left(\max(\mathbf{u}_i) - \min(\mathbf{u}_i)\right)}$$

Muestra	$\mathbf{u}_i$	$\mathbf{u}_i - \min(\mathbf{u}_i)$	$\frac{\left(\mathbf{u}_i - \min(\mathbf{u}_i)\right)}{\left(\max(\mathbf{u}_i) - \min(\mathbf{u}_i)\right)}$	$\mathbf{v}_i$
muestra 1	4	2	2/8	0.25
muestra 2	6	4	4/8	0.50
muestra 3	8	6	6/8	0.75
muestra 4	2	0	0/8	0
muestra 5	10	8	8/8	1

- Técnicas de normalización

Min-Max

Media desviación

- El método de normalización *media-desviación* transforma el rango de valores originales a media cero y desviación uno

- Media cero, desviación uno

$\mu = 0 \quad \sigma = 1$

$$\mathbf{v}_i = f_N(\mathbf{u}_i) = \frac{\mathbf{u}_i - \mu}{\sigma_i}$$

Muestra	$\mathbf{u}_i$	$\mathbf{u}_i - \mu$	$\mathbf{v}_i$
muestra 1	4	-2	-0.63
muestra 2	6	0	0
muestra 3	8	2	0.63
muestra 4	2	-4	-1.26
muestra 5	10	4	1.26

$$\sigma = \sqrt{\frac{(4-6)^2 + (6-6)^2 + (8-6)^2 + (2-6)^2 + (10-6)^2}{4}} = 3.16$$

Media

centro de gravedad de la muestra .

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Desviación Estándar

Es una medida de dispersión que nos dice cuan alejados estamos de la media

$$\sigma = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \mu)^2}$$

- Supervisados
  - kNN
  - Mahalanobis
  - Linear Discriminant Analysis
  - Árboles de decisión
  - Redes de Bayes
  - Redes Neuronales
- No supervisados
  - Mean Shift
  - K-Means
  - GMM-EM
  - Clustering jerárquico
  - Redes autoasociativas

OBJETIVO:
Definir los rangos de aceptación y rechazo para cada clase a partir de sus mediciones

- Rendimiento de clasificación binaria

  - El rendimiento del clasificador permite evaluar el porcentaje de error que genera la clasificación. Para evaluar dicho rendimiento se construye la matriz de confusión.

es \ predicho	Sano	Enfermo
Sano	TN	FP
Enfermo	FN	TP

En esta matriz sólo tenemos dos clases pero pueden tener todas las que queramos

  - Descripción

True Negative (TN): Número de personas sanas clasificadas como sanas
True Positive (TP): Número de personas enfermas clasificadas con enfermas

False Positive (FP): Número de personas sanas clasificadas con enfermas
False Negative (FN): Número de personas enfermas clasificadas como sanas

Idealmente estos deben tender a cero

Rendimiento de clasificación binaria

predicho →
es ↓

	Negative	Positive
Negative	TN	FP
Positive	FN	TP

True Negative (TN): Número de personas sanas clasificadas como sanas
True Positive (TP): Número de personas enfermas clasificadas con enfermas

False Positive (FP): Número de personas sanas clasificadas con enfermas
False Negative (FN): Número de personas enfermas clasificadas como sanas

True Positive Rate o Recall	True Negative Rate
$TPR = \frac{TP}{TP+FN}$	$TNR = \frac{TN}{TN+FP}$
False Positive Rate (Error T.I)	**False Negative Rate (Error T.2)**
$FPR = \frac{FP}{TN+FP}$	$FNR = \frac{FN}{FN+TP}$

Sensibilidad	Precisión
$S_n = \frac{TP}{TP+FN}$	$P = \frac{TP}{FP+TP}$
Especificidad	**F-Score**
$S_p = \frac{TN}{FP+TN}$	$F = 2 \times \frac{TPR \times P}{TPR + P}$

Más medidas en http://en.wikipedia.org/wiki/Sensitivity_and_specificity

Rendimiento de clasificación binaria

- EJEMPLO: Suponga que usted diseño un software para determinar automáticamente si usted enfermo o sano a través de un algoritmo de procesamiento de imágenes.

es ↓ / predicho →

	Negative	Positive
Negative	TN	FP
Positive	FN	TP

- Luego de una posterior clasificación, estos fueron los resultados. De 100 personas sanas, el sistema predijo que 95 estaban sanas y 5 enfermas. De 80 personas enfermas, el sistema predijo que 10 estaban sanas y 70 enfermas

es ↓ / predicho →

	Sano	Enfermo	Σ
Sano	95	5	100
Enfermo	10	70	80

True Positive Rate o Recall	True Negative Rate
$TPR = \dfrac{TP}{TP+FN}$	$TNR = \dfrac{TN}{TN+FP}$
False Positive Rate (Error T.I)	**False Negative Rate (Error T.2)**
$FPR = \dfrac{FP}{TN+FP}$	$FNR = \dfrac{FN}{FN+TP}$

Sensibilidad	Precisión o PPV
$S_n = \dfrac{TP}{TP+FN}$	$P = \dfrac{TP}{FP+TP}$
Especificidad	**F-Score**
$S_p = \dfrac{TN}{FP+TN}$	$F = 2 \times \dfrac{TPR \times P}{TPR + P}$

Más medidas en http://en.wikipedia.org/wiki/Sensitivity_and_specificity

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Rendimiento de clasificación binaria

- Resultados

1. La sensibilidad es 87.5%.
2. La especificidad es 95%.
3. La precisión es 93%
4. El F-Score es 90.32%

El óptimo es 100%

5. El Error Tipo 1 o FPR es de un 5%
6. El Error tipo 2 o FNR es de un 12.5%

El óptimo es 0%

Existen muchas técnicas de clasificación de optimización que buscan lograr el óptimo. Sin embargo, no siempre podrá ser determinado. En dicho caso deberemos determinar nuestro propio objetivo.

- Rendimiento

True Positive Rate o Recall	True Negative Rate
$TPR = \frac{70}{80}$	$TNR = \frac{95}{100}$
False Positive Rate (Error T.I)	**False Negative Rate (Error T.2)**
$FPR = \frac{5}{100}$	$FNR = \frac{10}{80}$

Sensibilidad	Precisión
$S_n = \frac{70}{70+10}$	$P = \frac{70}{5+70} = 0.933$
Especificidad	**F-Score**
$S_p = \frac{95}{5+95}$	$F = 2 \times \frac{0.875 \times 0.933}{0.875 + 0.933}$

- Supervisados

  - Los clasificadores supervisados son aquellos donde se conoce a priori la clase de cada instancia y el número de clases de la muestra.

- kNN
- Mahalanobis

no-paramétricos

Los clasificadores no-paramétricos no requieren entrenamiento previo ya que las relaciones se generan a partir de los datos.

- Clasificador Lineal
- Árboles de decisión
- Redes de Bayes
- Redes Neuronales

paramétricos

Son aquellos clasificadores que requieren un entrenamiento previo para generar una función de transferencia.

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

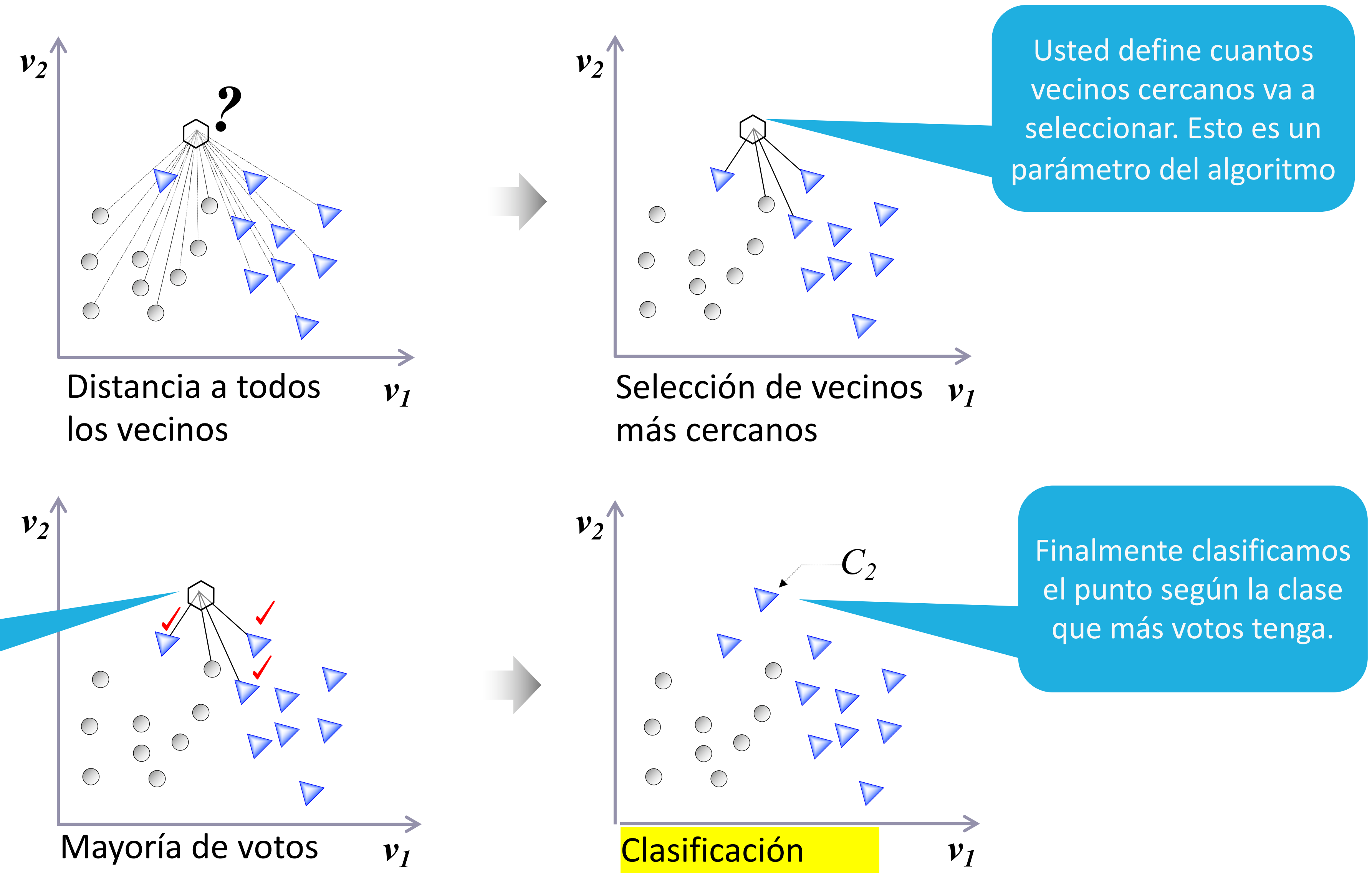
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- El clasificador *k*-Vecinos Cercanos (*k*-NN) calcula la distancia de los *k* vecinos más cercanos al punto. Normalmente se utiliza en problemas con baja dimensión (menor a 20) y con 8 vecinos.

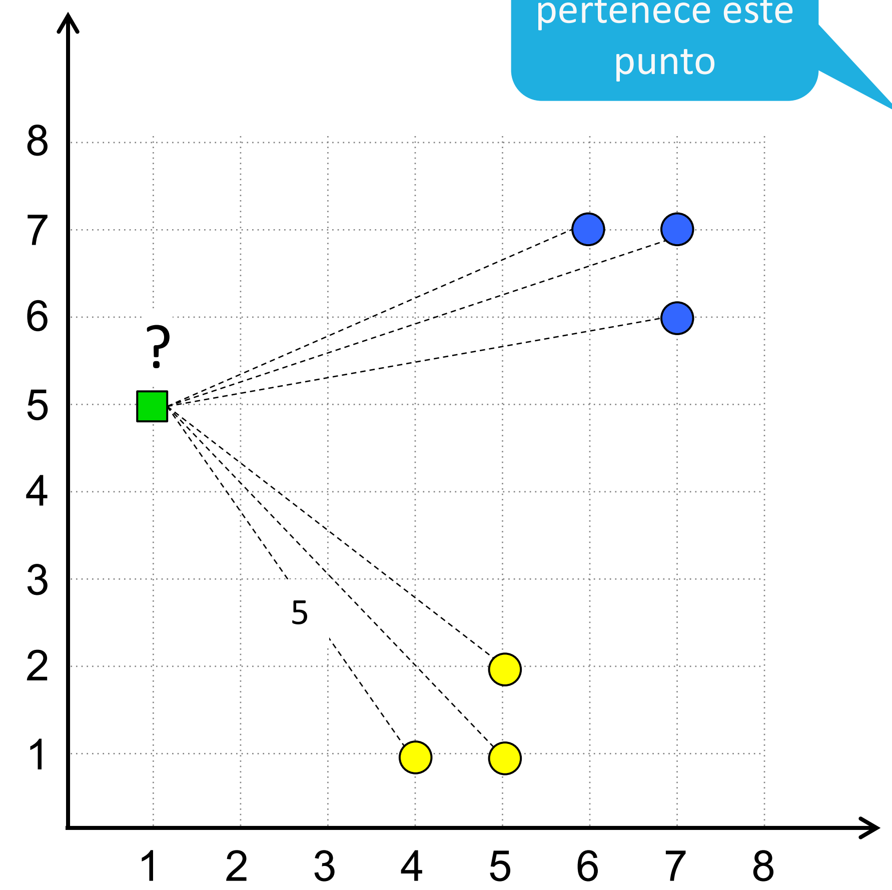


- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO: Calcular la distancia del punto que deseamos conocer su clase respecto a todos los puntos del problema.

A qué clase pertenece este punto



x	*y*
1	5

x	*y*	*clase*	*distancia*
4	1	1	5.00
5	1	1	5.66
5	2	1	5.00
6	7	2	5.39
7	6	2	6.08
7	7	2	6.32

Podemos ocupar cualquier métrica de distancia

$$\sqrt{(4-1)^2+(1-5)^2}=5$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

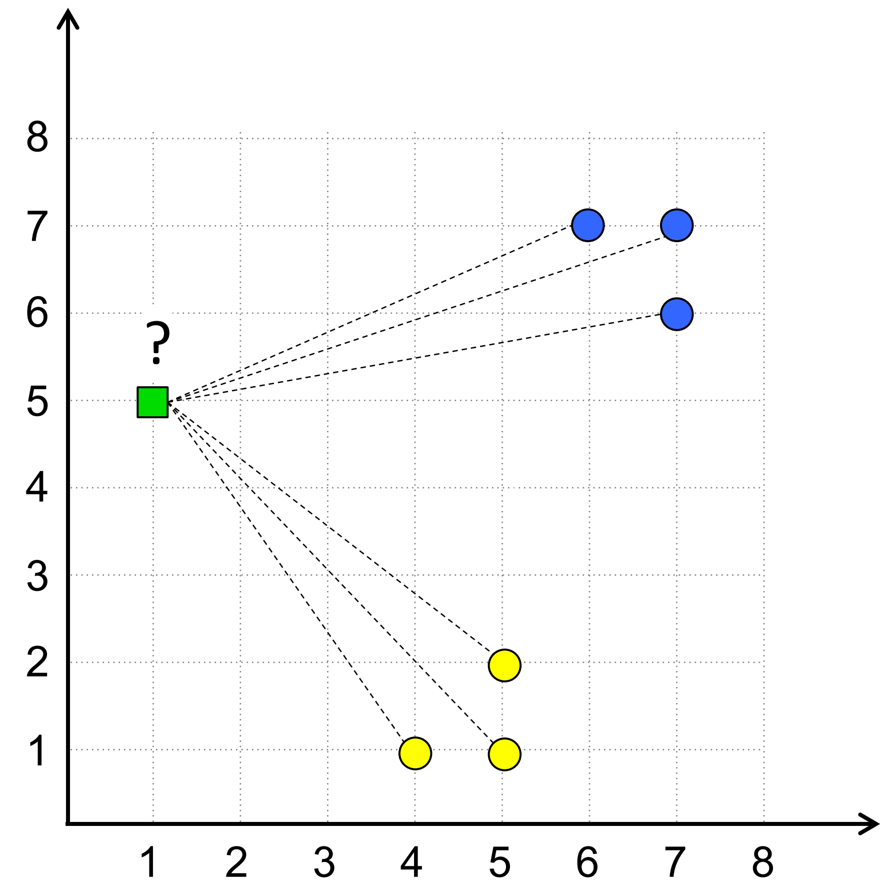
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- SEGUNDO: Ordenamos la distancia en orden ascendente y seleccionamos los k más cercanos, donde k es un número natural mayor o igual a 1.



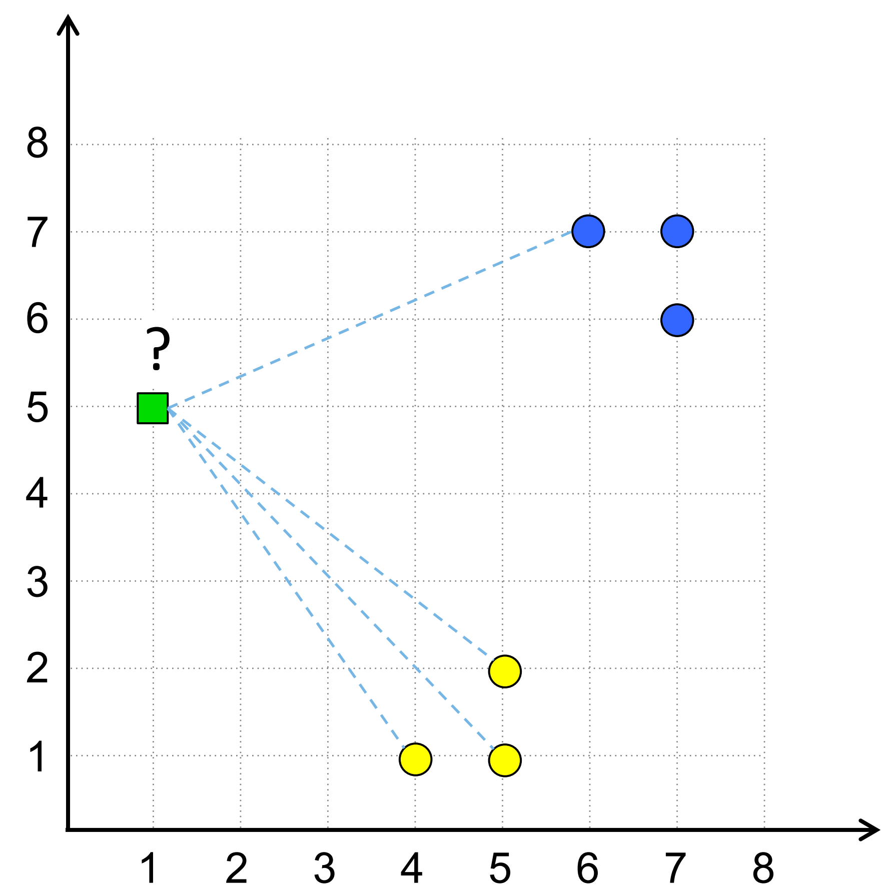
distancia	*clase*
5.00	1
5.66	1
5.00	1
5.39	2
6.08	2
6.32	2

distancia ordenada	*clase*
5.00	1
5.00	1
5.39	2
5.66	1
6.08	2
6.32	2

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- TERCERO: de la lista de puntos más cercanos, extraemos la clase de los k vecinos más cercanos.



distancia ordenada	*clase*
5.00	1
5.00	1
5.39	2
5.66	1
6.08	2
6.32	2

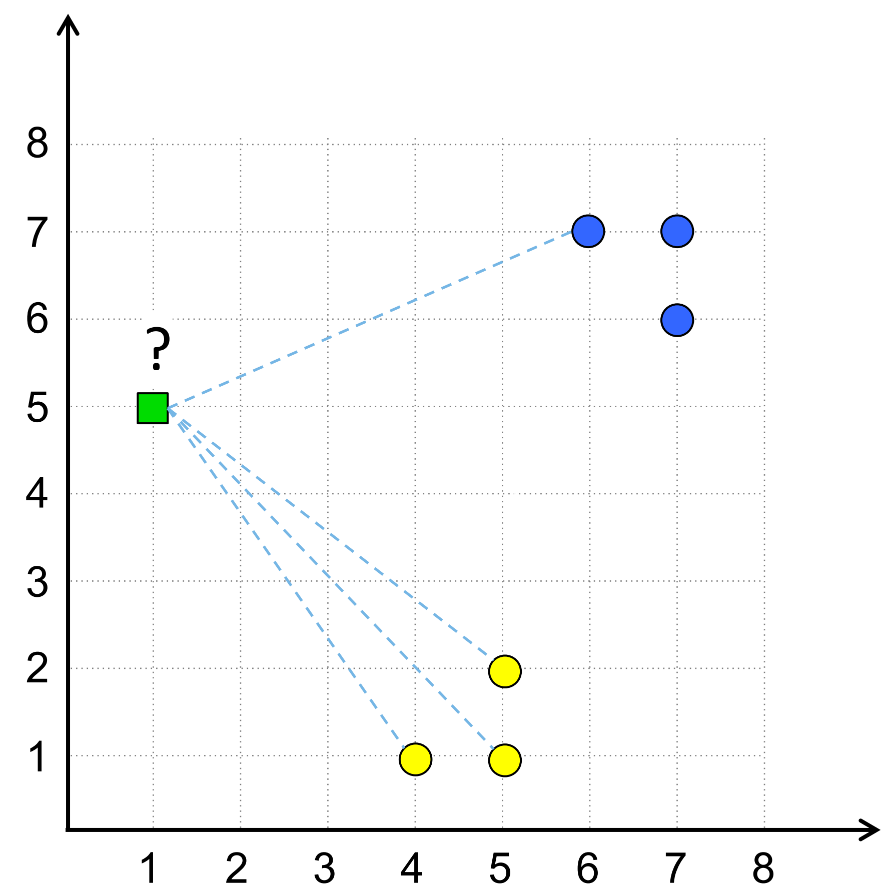
k-clase
1
1
2
1

Resultado de las cuatro clases más cercanas

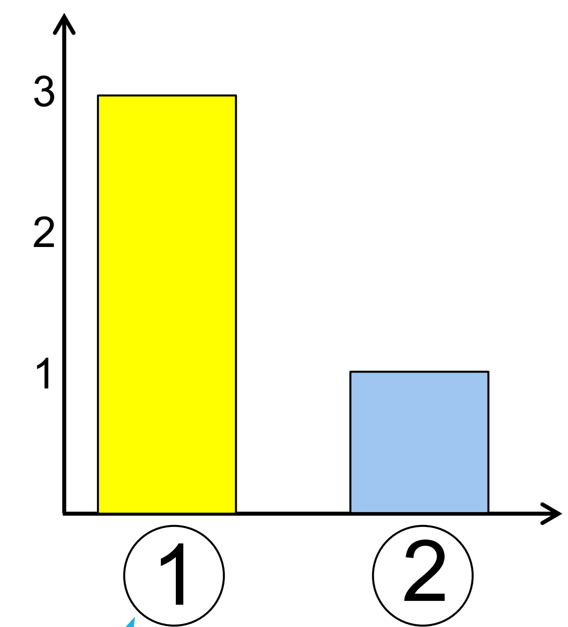
Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- CUARTO: Realizamos la mayoría de votos según cada clase que se encuentre en dicha lista y seleccionamos la clase que tenga más votos.



distancia ordenada	*clase*
5.00	1
5.00	1
5.39	2
5.66	1
6.08	2
6.32	2



En el ejemplo vamos determinamos el número de clases de los 4 vecinos más cercanos

FACULTAD DE INGENIERÍA Y CIENCIAS

- Supervisados

kNN

Mahalanobis

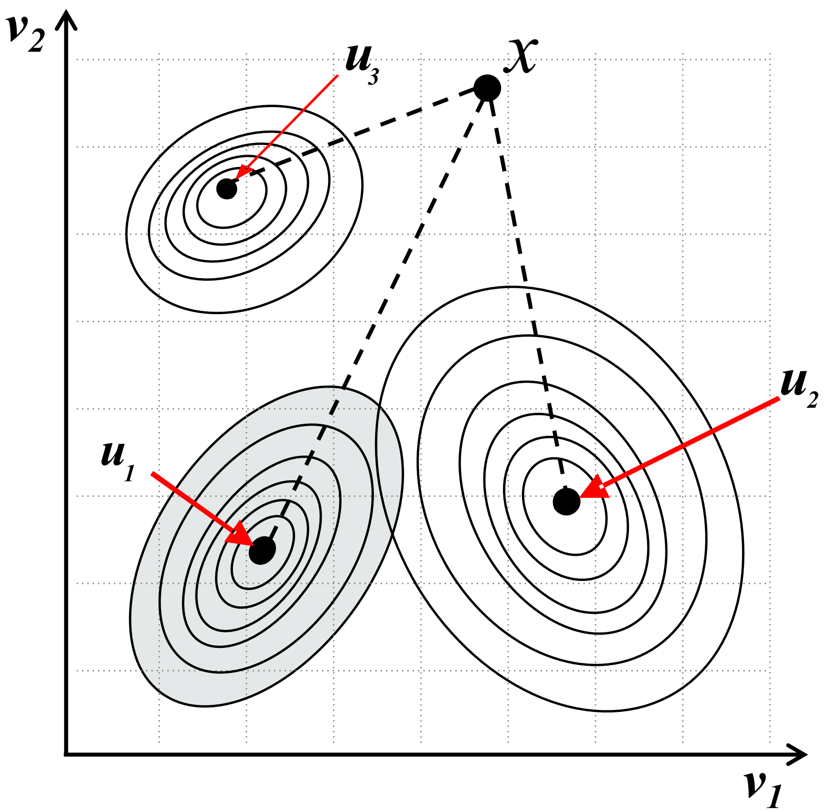
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- La métrica de distancia de Mahalanobis mide la similitud entre dos variables aleatorias multidimensionales.



$$d_1 = \left(x - \bar{x}_1\right)^{T} \Sigma_1^{-1} \left(x - \bar{x}_1\right)$$

Centro de masa de la clase 1.

Matriz inversa de covarianza de la clase 1

- A diferencia de la distancia euclidiana, la distancia de Mahalanobis mide la dependencia de las variables a través de la matriz de correlación, incorporando así una ponderación de las distintas escalas de cada variable.

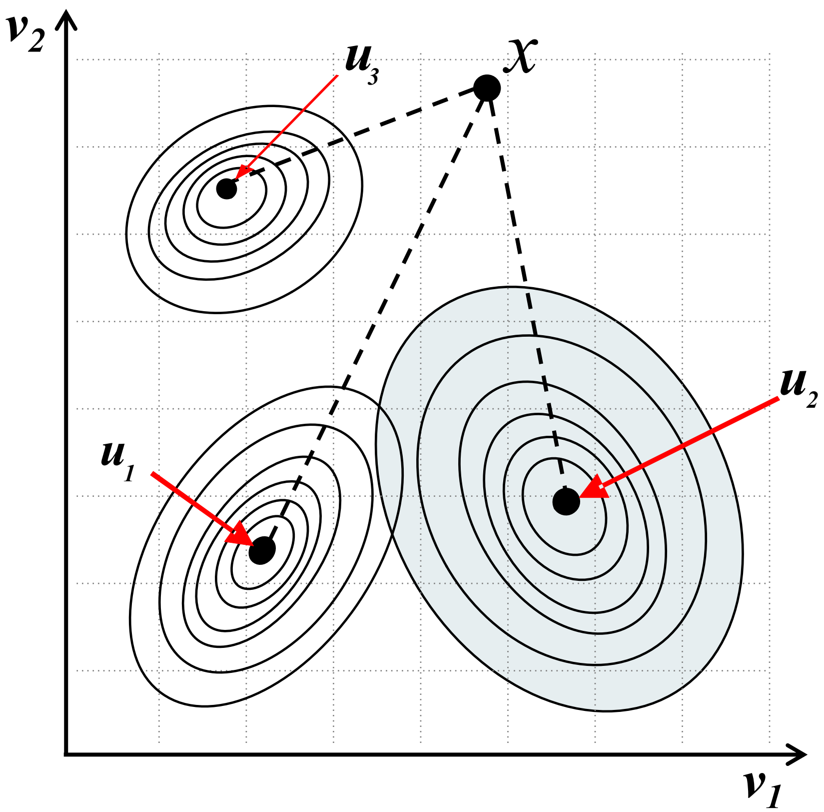
FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- La métrica de distancia de Mahalanobis mide la similitud entre dos variables aleatorias multidimensionales.



$$d_1 = \left(x - \bar{x}_1\right)^{\mathrm{T}} \Sigma_1^{-1} \left(x - \bar{x}_1\right)$$

$$d_2 = \left(x - \bar{x}_2\right)^{\mathrm{T}} \Sigma_2^{-1} \left(x - \bar{x}_2\right)$$

Centro de masa de la clase 2.

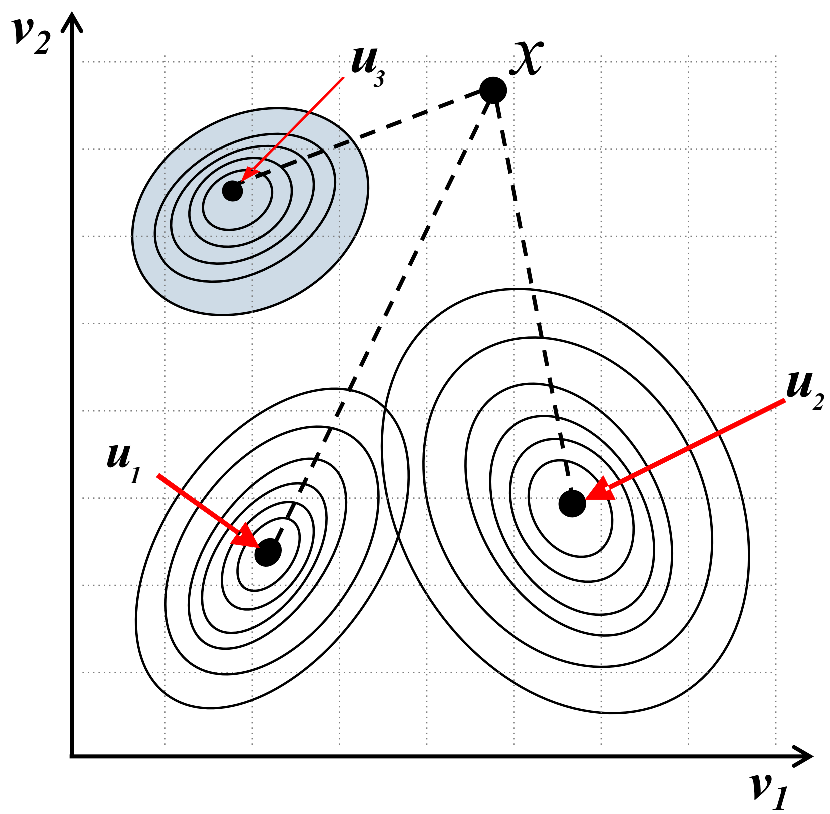
Matriz inversa de covarianza de la clase 1

- A diferencia de la distancia euclidiana, la distancia de Mahalanobis mide la dependencia de las variables a través de la matriz de correlación, incorporando así una ponderación de las distintas escalas de cada variable.

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- La métrica de distancia de Mahalanobis mide la similitud entre dos variables aleatorias multidimensionales.



$$d_1 = (x - \bar{x}_1)^T \Sigma_1^{-1} (x - \bar{x}_1)$$

$$d_2 = (x - \bar{x}_2)^T \Sigma_2^{-1} (x - \bar{x}_2)$$

$$d_3 = (x - \bar{x}_3)^T \Sigma_3^{-1} (x - \bar{x}_3)$$

Centro de masa de la clase 3.

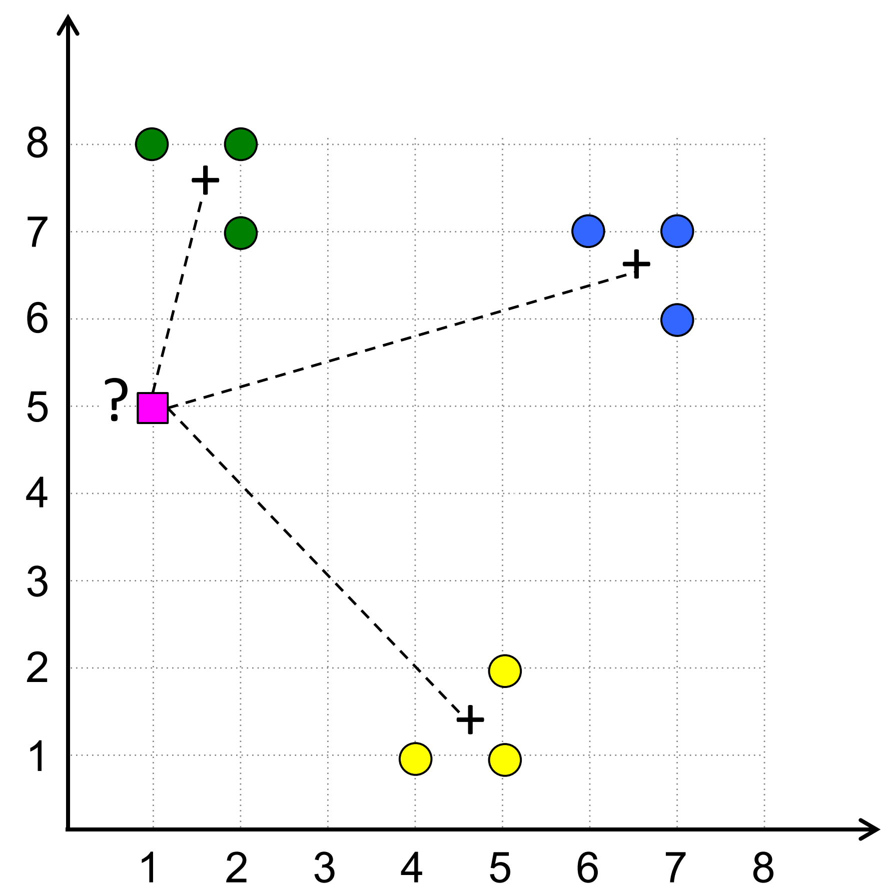
Matriz inversa de covarianza de la clase 1

- A diferencia de la distancia euclidiana, la distancia de Mahalanobis mide la dependencia de las variables a través de la matriz de correlación, incorporando así una ponderación de las distintas escalas de cada variable.

Supervisados

- kNN
- **Mahalanobis**
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- La métrica de distancia de Mahalanobis mide la similitud entre dos variables aleatorias multidimensionales. Veamos un ejemplo para clasificar un nuevo punto



A qué clase pertenece este punto

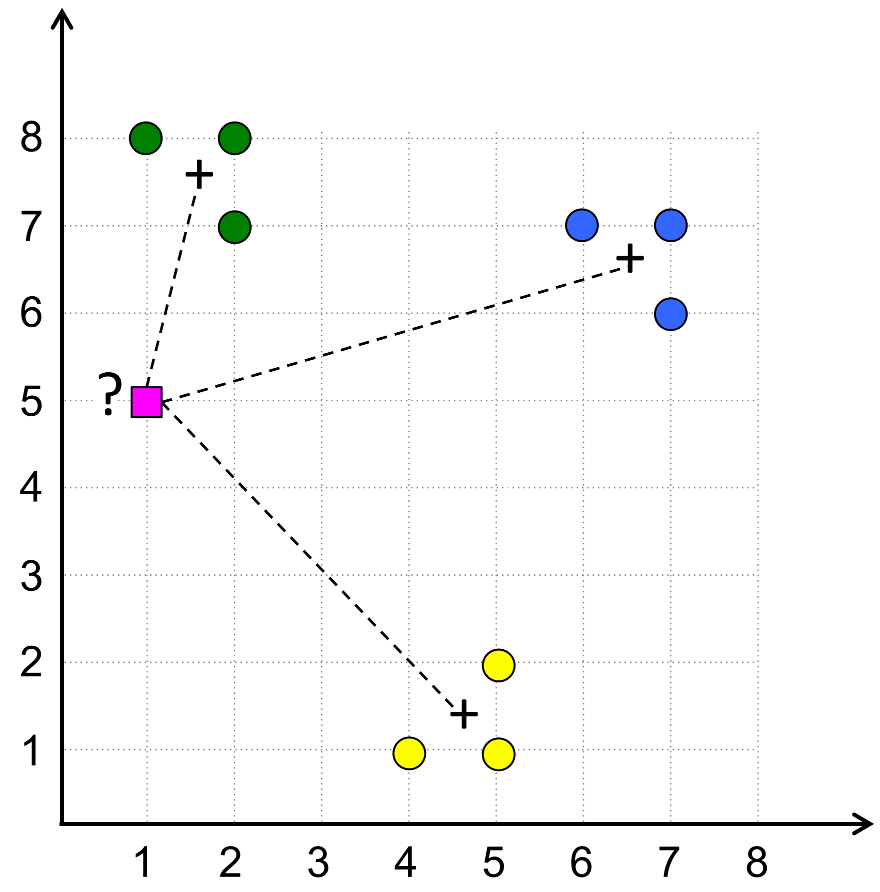
x	y
1	5

x	y	*clase*
4	1	1
5	1	1
5	2	1
6	7	2
7	6	2
7	7	2
1	8	3
2	7	3
2	8	3

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO: Calcular la media de cada clase. Para ello debemos separar cada clase del conjunto de datos.



x	y
1	5

x	y
4	1
5	1
5	2

mu_1 → 4.6 1.3

x	y
6	7
7	6
7	7

mu_2 → 6.6 6.6

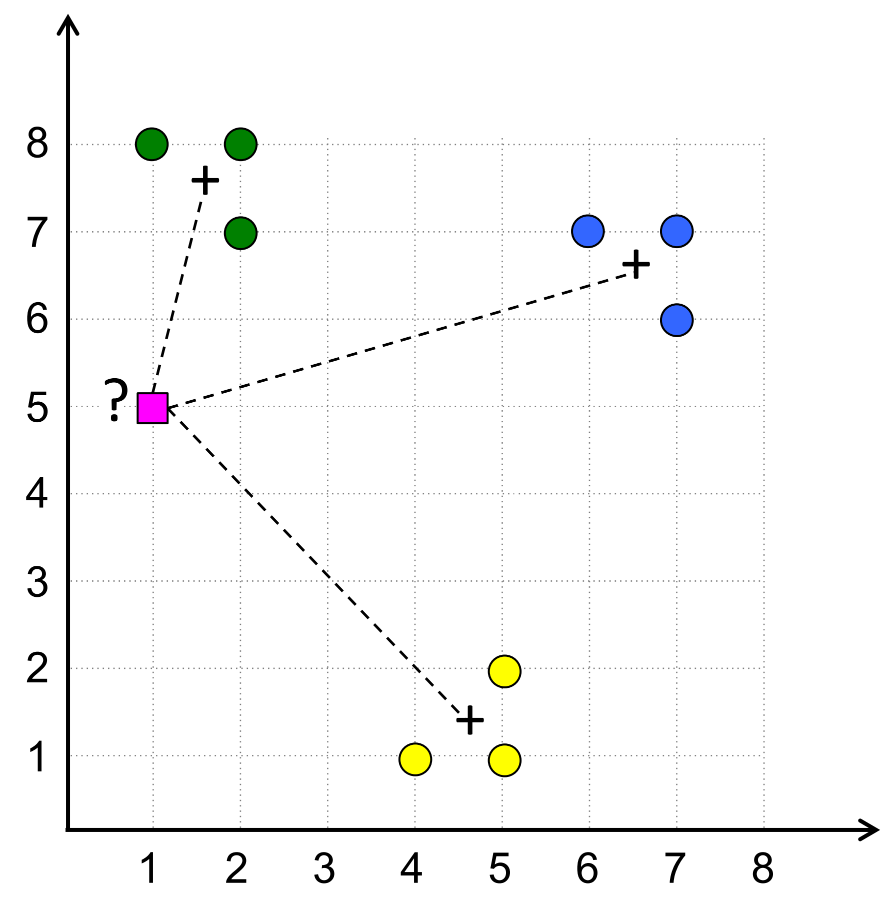
x	y
1	8
2	7
2	8

mu_3 → 1.6 7.6

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- SEGUNDO: Calcular la covarianza de cada clase. Recuerde aplicar la covarianza sobre los datos por clase y no al conjunto completo de datos.



x	y
1	5

x	y
4	1
5	1
5	2

x	y
6	7
7	6
7	7

x	y
1	8
2	7
2	8

$$\mathrm{cov}_1 = \begin{bmatrix} 0.33 & 0.16 \\ 0.16 & 0.33 \end{bmatrix}$$

$$\mathrm{cov}_2 = \begin{bmatrix} 0.33 & -0.16 \\ -0.16 & 0.33 \end{bmatrix}$$

$$\mathrm{cov}_3 = \begin{bmatrix} 0.33 & -0.16 \\ -0.16 & 0.33 \end{bmatrix}$$

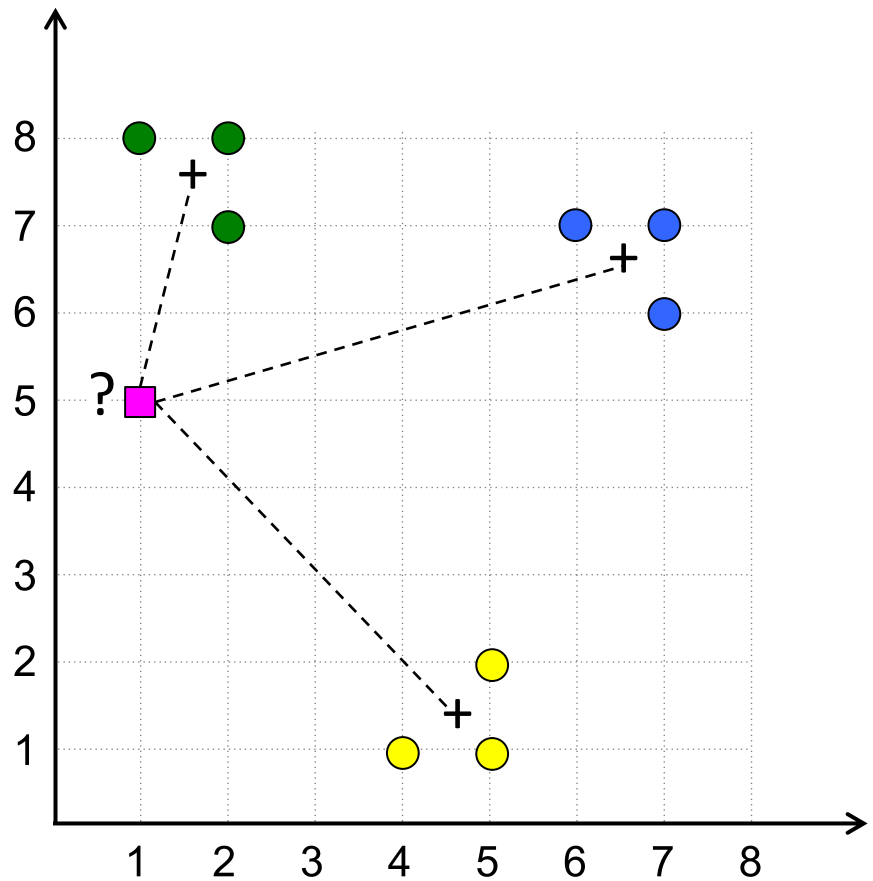
FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- **Mahalanobis**
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- TERCERO: Por cada clase, determinar la distancia de Mahalanobis a cada una de las clases empleando la media y la covarianza de cada clase.



$$d = \left(x - \bar{x}\right)^{T} \Sigma_1^{-1} \left(x - \bar{x}\right)$$

x	y
1	5

x	y
4	1
5	1
5	2

x	y
6	7
7	6
7	7

x	y
1	8
2	7
2	8

$$\begin{bmatrix}-3.6 & 3.6\end{bmatrix}\begin{bmatrix}0.33 & 0.16\\ 0.16 & 0.33\end{bmatrix}^{-1}\begin{bmatrix}-3.6\\ 3.6\end{bmatrix} \rightarrow d = 161.33$$

$$\begin{bmatrix}-5.6 & -1.6\end{bmatrix}\begin{bmatrix}0.33 & -0.16\\ -0.16 & 0.33\end{bmatrix}^{-1}\begin{bmatrix}-5.6\\ -1.6\end{bmatrix} \rightarrow d = 177.33$$

$$\begin{bmatrix}-0.6 & 2.6\end{bmatrix}\begin{bmatrix}0.33 & -0.16\\ -0.16 & 0.33\end{bmatrix}^{-1}\begin{bmatrix}-0.6\\ 2.6\end{bmatrix} \rightarrow d = 37.33$$

- Supervisados

kNN

Mahalanobis

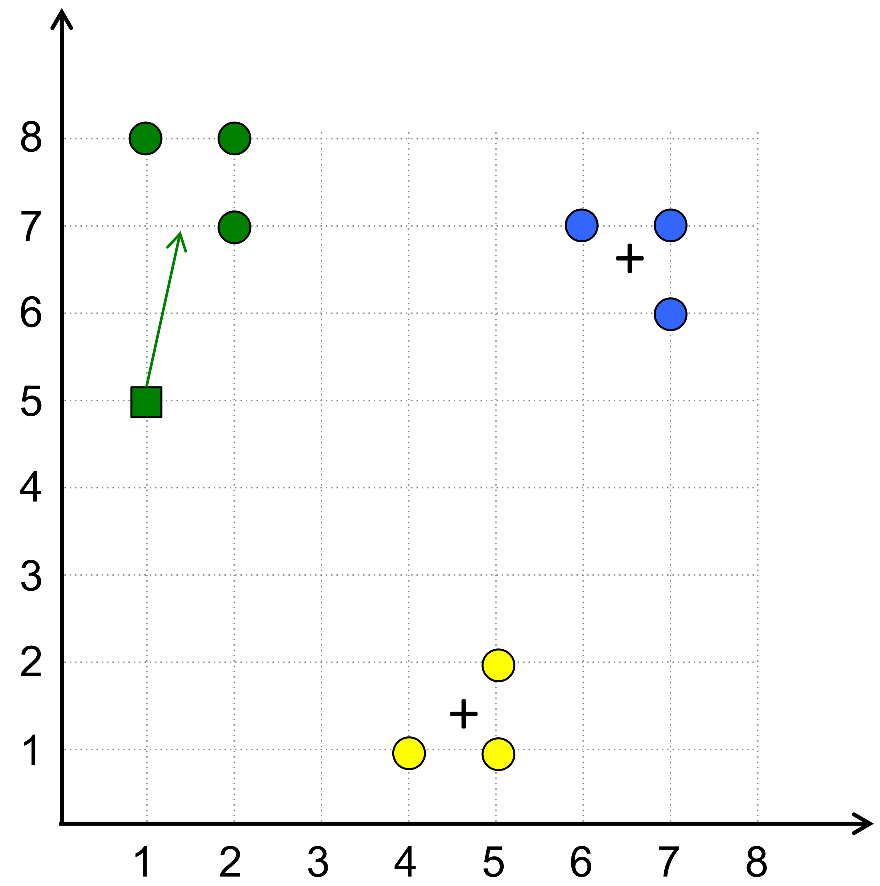
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- CUARTO: Asignamos aquella clase que posea la menor distancia de Mahalanobis entre todas las clases o clusters



x	y
1	5

x	y
4	1
5	1
5	2

$d = 161.33$

x	y
6	7
7	6
7	7

$d = 177.33$

x	y
1	8
2	7
2	8

$d = 37.33$

La menor distancia se encuentra en la clase 3

- Supervisados

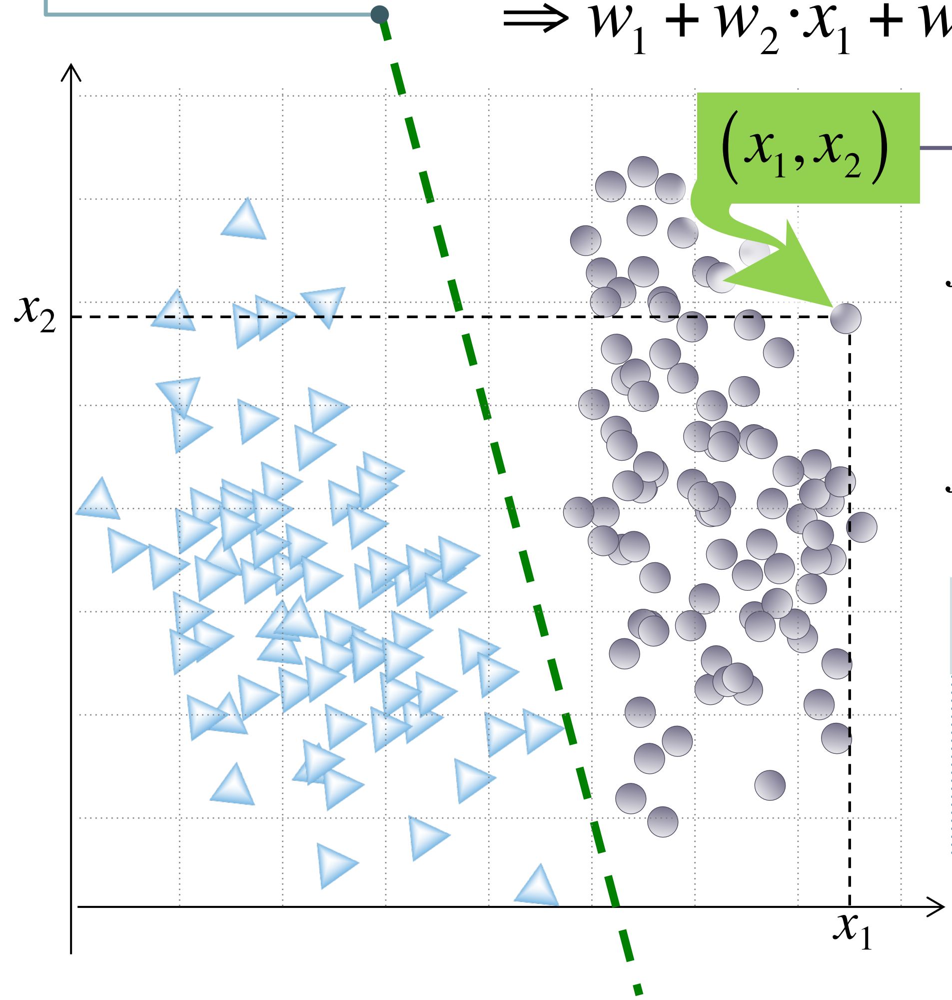
kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- El clasificador lineal corresponde a una línea de separación entre dos clases. El problema consiste en cómo encontrar los parámetros que definen dicha línea.

PROBLEMA: ¿Cómo encontrar el vector W?

$$x_2 = m\cdot x_1 + c \Rightarrow c + m\cdot x_1 - x_2 = 0$$

$$\Rightarrow w_1 + w_2\cdot x_1 + w_3\cdot x_2 = 0$$



$$f(x_1, x_2) = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = 0$$

$$f(x_1, x_2) = w_0 + x_1 w_1 + x_2 w_2 = 0$$

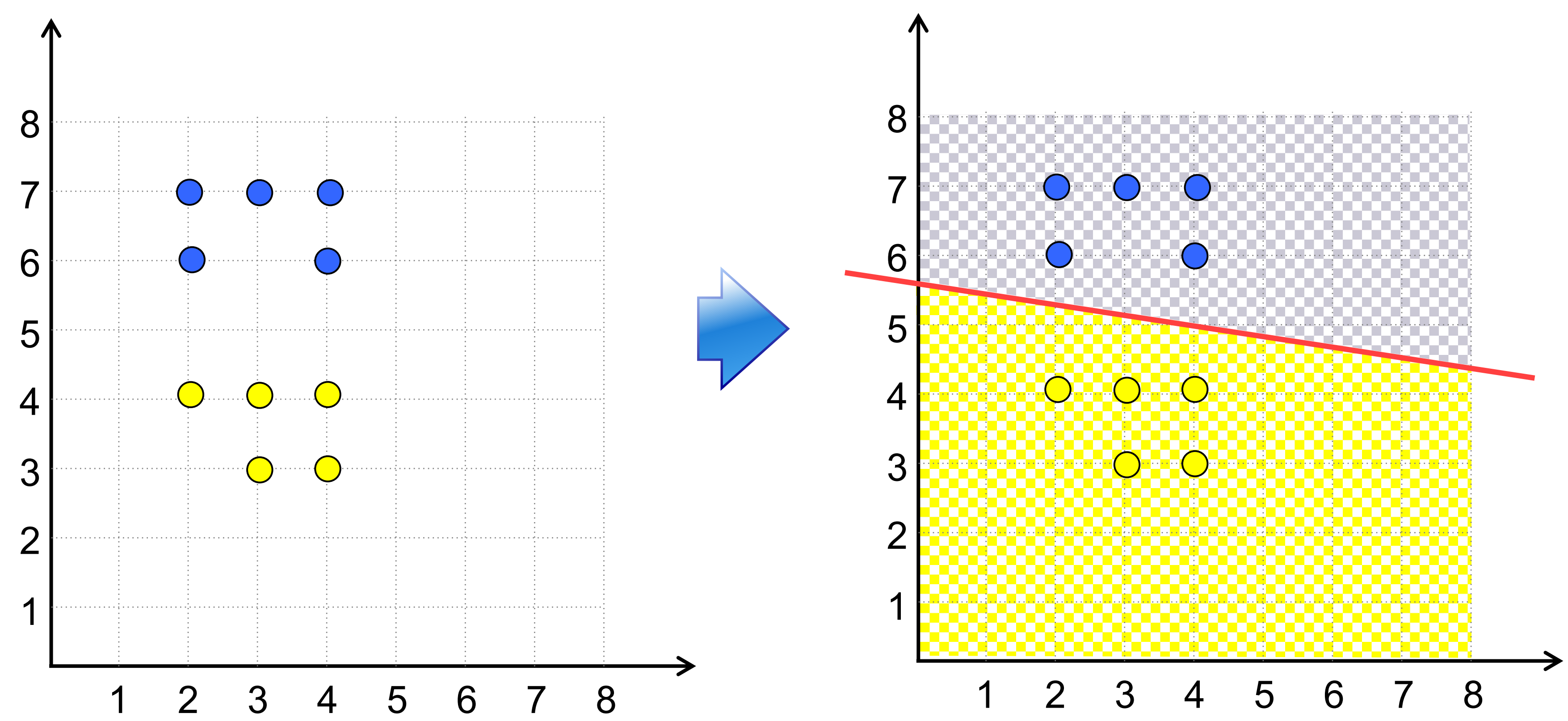
Clasificación	
si	$f > 0 \rightarrow$ clase 0
si	$f < 0 \rightarrow$ clase 1

FACULTAD DE INGENIERÍA Y CIENCIAS
UAI UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- *Linear Discriminant Analysis*
  Es una técnica directa para encontrar los parámetros de la línea de decisión y al mismo tiempo permite determinar la clasificación de los puntos



FACULTAD DE INGENIERÍA Y CIENCIAS

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano

- Definiciones:
  - En un problema de múltiples clases, la probabilidad a priori de la clase k es π_k donde

$$\sum_{k=1}^{K} \pi_k = 1$$

  - π_k es usualmente estimado en forma empírica calculando las frecuencias del set de entrenamiento

$$\hat{\pi}_k = \frac{\#\text{ de ejemplos clase } k}{\#\text{ total de ejemplos}}$$

Objetivo: ¿Cual es la probabilidad de pertenecer a la clase k dado el conjunto de medidas x?

En otras palabras, estimar la probabilidad a posteriori de la clase k

$$\longrightarrow P(k \mid \mathbf{x})$$

Clase k-ésimo

Dato

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Demostración:**
  - Empleando el teorema de Bayes obtenemos

función multivariada

$$P(k \mid \mathbf{x}) = \frac{f_k(\mathbf{x}) \cdot \pi_k}{\sum_{l=1}^{K} f_l(\mathbf{x}) \cdot \pi_l}$$

  - Dado que solo buscamos el máximo, descartamos el denominador. Luego buscamos el máximo de todas las clases (recuerde que tenemos k clases). Por lo tanto, el máximo corresponde a:

$$\begin{aligned}\hat{G}(\mathbf{x}) &= \arg\max_k \left\{ P(k \mid \mathbf{x}) \right\} \\ &= \arg\max_k \left\{ f_k(\mathbf{x}) \cdot \pi_k \right\} \\ &= \arg\max_k \left\{ \log\left(f_k(\mathbf{x}) \cdot \pi_k \right) \right\} \end{aligned} \quad (1)$$

Gaussiana Multivariada

covarianza interclases

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^{\mathrm{T}} \Sigma^{-1} (\mathbf{x}-\mu_k)} \quad (2)$$

media de la clase k

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **Demostración:**
  - Reemplazando (2) en (1) tenemos

$$\hat{G}(\mathbf{x}) = \arg\max_k \left\{ \log\left(\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu_k)} \cdot \pi_k \right) \right\}$$

  - Evaluando el logaritmo y expandiendo términos

$$\begin{aligned}
\hat{G}(\mathbf{x}) &= \arg\max_k \left\{ -\frac{1}{2}(\mathbf{x}-\mu_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu_k) + \log\left(\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\right) + \log(\pi_k) \right\} \\
&= \arg\max_k \left\{ \underbrace{-\frac{1}{2}(\mathbf{x}-\mu_k)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu_k)} - \log\left((2\pi)^{p/2}|\Sigma|^{1/2}\right) + \log(\pi_k) \right\} \\
&= \arg\max_k \left\{ \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k \cancel{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}} \cancel{-\log\left((2\pi)^{p/2}|\Sigma|^{1/2}\right)} + \log(\pi_k) \right\}
\end{aligned}$$

  - Eliminando las constantes, el máximo corresponde a:

$$\hat{G}(\mathbf{x}) = \arg\max_k \left\{ \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k) \right\}$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

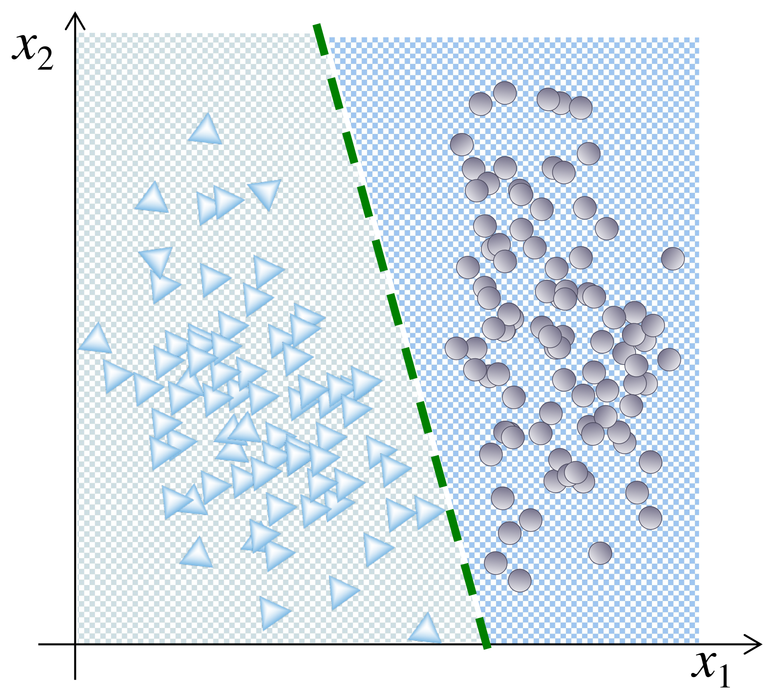
- Clasificación:
  - La función lineal de decisión es

Función LDA

$$f_k(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k)$$

  - De esta forma para asignar una clase a un punto $\boldsymbol{x}$ solo debemos calcular la función $f_k(\mathbf{x})$ que sea máxima.

$$f(\mathbf{x}) = \arg\max_k \{\delta_k(\mathbf{x})\}$$



  - Recordemos que nuestro problema original consiste en buscar los parámetros de la línea de decisión.

DATOS

$$f(x_1, x_2) = w_0 + x_1 w_1 + x_2 w_2 = 0$$

INCÓGNITAS

- Supervisados

kNN

Mahalanobis

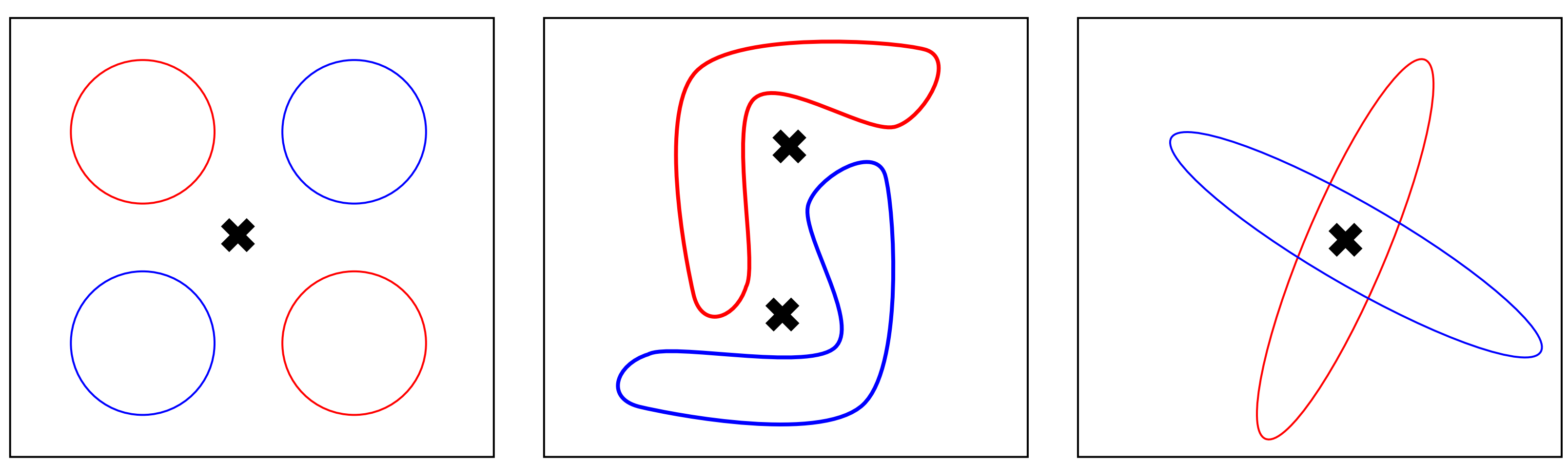
LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **Lo bueno**
  - LDA modela las diferencias entre las clases.
  - Funciona bien si las distribuciones son normales (Gaussianas).
  - Si existe una combinación lineal de las características, entonces LDA obtendrá un buen rendimiento.
  - LDA es conocido por reducir la dimensionalidad. Así la proyección de múltiples dimensiones es reducida a un espacio de menor dimensión
- **Lo malo**
  - Si las distribuciones son significativamente no gaussianas, las proyecciones de LDA no podrán preservar la compleja estructura de los datos.



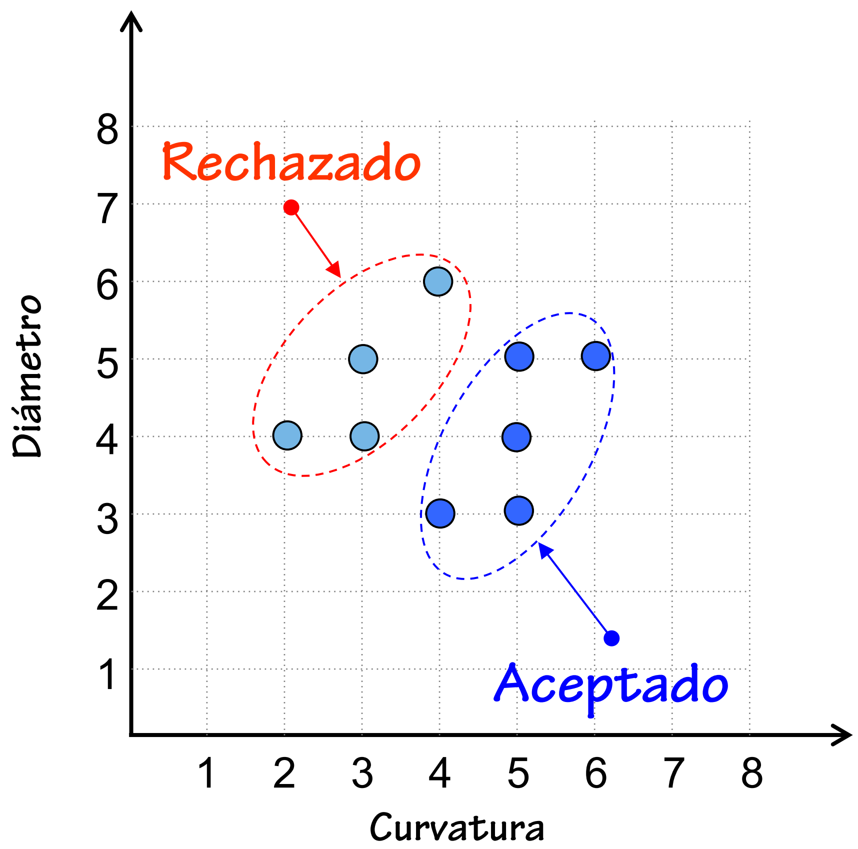
FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- Mahalanobis
- LDA
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **Ejemplo:**
  Supongamos que la empresa *Mckoy* produce galletas de muy alta calidad, por ello tiene un muy exhaustivo control de calidad.

  La empresa desea automatizar su sistema para que automáticamente determine el rango de aceptación según las características extraídas.



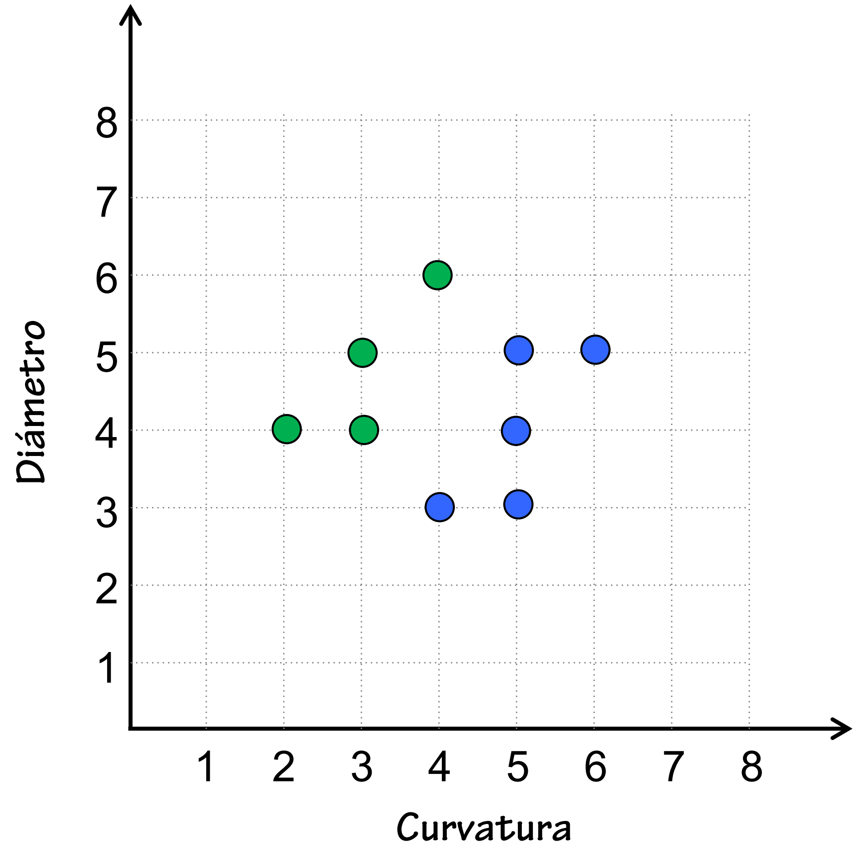
Curvatura	Diámetro	Control Calidad
4	3	Aceptado
5	3	Aceptado
5	4	Aceptado
5	5	Aceptado
6	5	Aceptado
2	4	Rechazado
3	4	Rechazado
3	5	Rechazado
4	6	Rechazado

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- **Preliminar:**
Separemos los datos en sus correspondientes clases y calculemos la media total

#	Curvatura	Diámetro	Clase
1	4	3	1
2	5	3	1
3	5	4	1
4	5	5	1
5	6	5	1
6	2	4	2
7	3	4	2
8	3	5	2
9	4	6	2



- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- PRIMERO:
  Calcule la probabilidad de cada clase según el número de elementos por clase

#	Curvatura	Diámetro	Clase
1	4	3	1
2	5	3	1
3	5	4	1
4	5	5	1
5	6	5	1
6	2	4	2
7	3	4	2
8	3	5	2
9	4	6	2

$$\hat{\pi}_k = \frac{\#\text{ de ejemplos clase } k}{\#\text{ total de ejemplos}}$$

Probabilidad a Priori
$\pi_1 = \frac{5}{9}, \quad \pi_2 = \frac{4}{9}$

Media	
4.11	4.33

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **SEGUNDO:**

  Calcule la covarianza y la media de clada clase por separado

DATOS EN CLASES	MEDIAS	COVARIANZAS

X_1

4	3
5	3
5	4
5	5
6	5

X_2

2	4
3	4
3	5
4	6

$\overline{x}_1$	5.0	4.0
$\overline{x}_2$	3.0	4.8

Cov(X_1)

0.500	0.500
0.500	1.000

Cov(X_2)

0.667	0.667
0.668	0.917

Recuerde:

$$\operatorname{cov}(x,y) = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- TERCERO:
  Calcule la matriz de covarianza entre clases ponderando la probabilidad de cada clase.

$Cov(X_1)$	
0.500	0.500
0.500	1.000

$Cov(X_2)$	
0.667	0.667
0.668	0.917

$$\Sigma = \pi_1 \cdot \mathrm{cov}\left(\mathbf{X}_1\right) + \pi_2 \cdot \mathrm{cov}\left(\mathbf{X}_2\right)$$

Σ	
0.5741	0.5741
0.5741	0.9630

En este ejemplo tenemos dos clases, por ello la suma solo utiliza la covarianza entre dos grupos

Σ^{-1}	
4.3134	-2.5714
-2.5714	2.5714

Esta matriz será luego empleada para determinar la clasificación lineal

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- **CUARTO:**
  A cada punto original, calculamos la función de decisión lineal.

$$f_k(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k)$$

#	Curvatura	Diámetro	f_1	f_2	Max(f_1,f_2)
1	4	3	13.76	3.82	1
2	5	3			
3	5	4			
4	5	5			
5	6	5			
6	2	4			
7	3	4			
8	3	5			
9	4	6			

$$f_1(\mathbf{x}) = \begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 5 & 4 \end{bmatrix}\begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} + \log\left(\frac{5}{9}\right) = 13.76$$

$$f_2(\mathbf{x}) = \begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\begin{bmatrix} 3 \\ 4.8 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 3 & 4.8 \end{bmatrix}\begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\begin{bmatrix} 3 \\ 4.8 \end{bmatrix} + \log\left(\frac{4}{9}\right) = 3.82$$

FACULTAD DE INGENIERÍA Y CIENCIAS

UAI UNIVERSIDAD ADOLFO IBÁÑEZ

Supervisados

- kNN
- Mahalanobis
- **LDA**
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- CUARTO:
  A cada punto original, calculamos la función de decisión lineal.

#	Curvatura	Diámetro	f_1	f_2	Clasificación
1	4	3	13.76	3.82	1
2	5	3	25.04	4.54	1
3	5	4	22.47	9.04	1
4	5	5	19.90	13.54	1
5	6	5	31.18	14.27	1
6	2	4	-11.37	6.86	2
7	3	4	-0.09	7.59	2
8	3	5	-2.66	12.09	2
9	4	6	6.05	17.31	2

Clase original
1
1
1
1
1
2
2
2
2

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

Redes Neuronales

- Parámetros de la recta:
  - Si queremos determinar los parámetros de la recta, notemos que se debe cumplir la siguiente condición. Es decir, buscamos la línea de decisión entre la clase k y la clase p

$$f_k(\mathbf{x}) = f_p(\mathbf{x})$$

$$f_k(\mathbf{x}) - f_p(\mathbf{x}) = 0$$

$$\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k) - \mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mu_p + \frac{1}{2}\mu_p^{\mathrm{T}}\Sigma^{-1}\mu_p - \log(\pi_p) = 0$$

  - Agrupando términos obtenemos

$$\mathbf{x}^{\mathrm{T}}\Sigma^{-1}(\mu_k - \mu_p) - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \frac{1}{2}\mu_p^{\mathrm{T}}\Sigma^{-1}\mu_p + \log\left(\frac{\pi_k}{\pi_p}\right) = 0$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Parámetros de la recta:
  - Finalmente la línea de decisión entre la clase k y la clase p es:

$$\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\left(\mu_k-\mu_p\right)-\frac{1}{2}\left(\mu_k+\mu_p\right)^{\mathrm{T}}\Sigma^{-1}\left(\mu_k-\mu_p\right)+\log\left(\frac{\pi_k}{\pi_p}\right)=0$$

  - Observe con detalle la relación entre estas dos ecuaciones ¿Que tienen en común?

Ecuación de la recta

$$w_0+x_1w_1+x_2w_2 \Rightarrow w_0+\mathbf{x}^{\mathrm{T}}\cdot\mathbf{w}=0$$

$$\mathbf{w}=-\Sigma^{-1}\left(\mu_k-\mu_p\right)$$

$$w_0=-\frac{1}{2}\left(\mu_k+\mu_p\right)^{\mathrm{T}}\Sigma^{-1}\left(\mu_k-\mu_p\right)+\log\left(\frac{\pi_k}{\pi_p}\right)$$

- Supervisados

kNN
Mahalanobis
LDA
Árboles de decisión
Bayesiano
Redes Neuronales

- Gráfico: Empleando las ecuaciones y valores previos, determinamos los parámetros de la recta

$$\mathbf{w} = \Sigma^{-1}\left(\mu_k - \mu_p\right) \qquad w_0 = -\frac{1}{2}\left(\mu_k - \mu_p\right)^{\mathrm{T}} \Sigma^{-1}\left(\mu_k - \mu_p\right) + \log\left(\frac{\pi_k}{\pi_p}\right)$$

$$\mathbf{w} = \begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\left(\begin{bmatrix} 5 & 4 \end{bmatrix} - \begin{bmatrix} 3 & 4.8 \end{bmatrix}\right) = \begin{bmatrix} 10.55 \\ -7.07 \end{bmatrix}$$

$$w_0 = -\frac{1}{2}\left(\begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 4.8 \end{bmatrix}\right)\begin{bmatrix} 4.313 & -2.571 \\ -2.571 & 2.571 \end{bmatrix}\left(\begin{bmatrix} 5 & 4 \end{bmatrix} - \begin{bmatrix} 3 & 4.8 \end{bmatrix}\right) + \log\left(\frac{5}{4}\right)$$

$$w_0 = -11.06$$

- Reemplazando los valores (forma general)

$$f\left(x_1, x_2\right) = w_0 + x_1 w_1 + x_2 w_2 = 0$$

$$f\left(x_1, x_2\right) = -11.06 + 10.55 x_1 - 7.07 x_2 = 0$$

FACULTAD DE INGENIERÍA Y CIENCIAS

- Supervisados

kNN

Mahalanobis

LDA

Árboles de decisión

Bayesiano

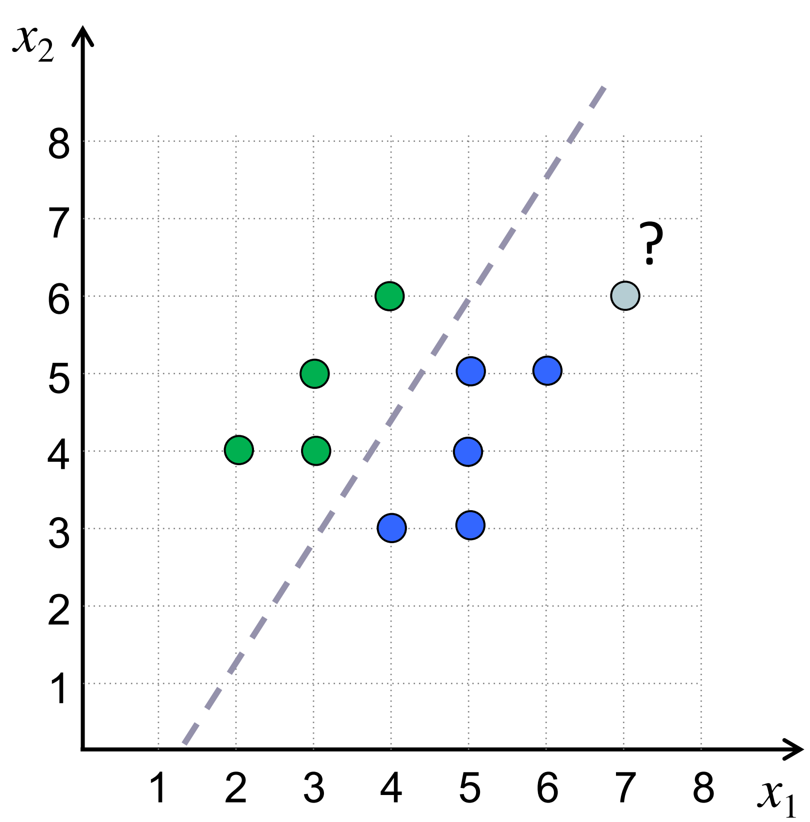
Redes Neuronales

- **Gráfico**

  Empleando las ecuaciones y valores previos, determinamos los parámetros de la recta

  - Ecuación de la recta general

$$f(x_1, x_2) = -11.06 + 10.55x_1 - 7.07x_2 = 0$$



$$x_2 = \frac{10.55}{7.07}x_1 - \frac{11.06}{7.07}$$

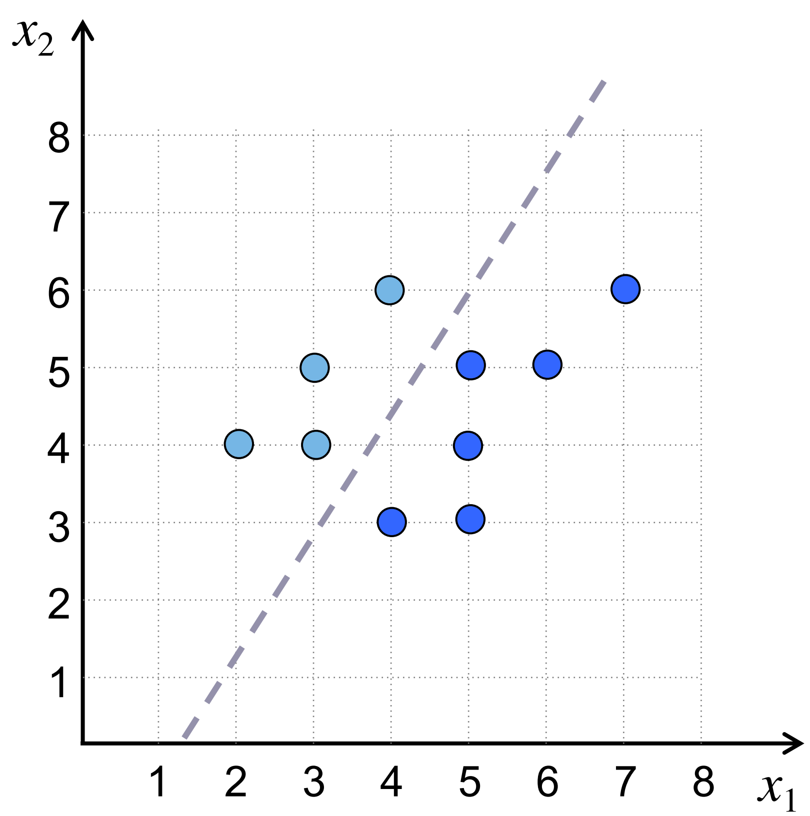
Curvatura

Supervisados

- kNN
- Mahalanobis
- **LDA**
- Árboles de decisión
- Bayesiano
- Redes Neuronales

- **Clasificación**
  Si tenemos un punto que deseamos evaluar si clase, simplemente verificamos a qué lado de la recta se encuentra. Veamos un ejemplo

  - Forma directa $x_2 = -\frac{10.55}{7.07}x_1 + \frac{11.06}{7.07}$



  - Evaluación en el punto $(7,6)$

$$x_2 = -\frac{10.55}{7.07} \times 7 + \frac{11.06}{7.07} = 8.88$$

$$8.88 > 6$$

$$entonces\ (7,6) \in$$