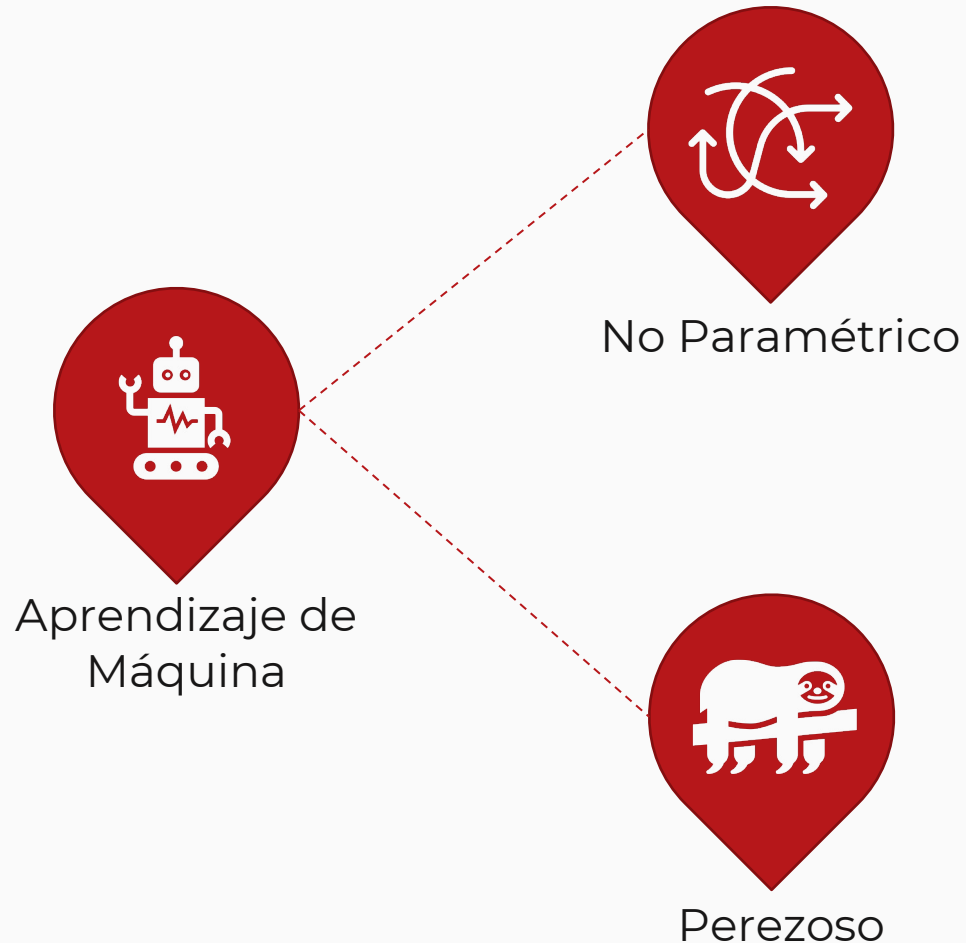


Otros Métodos de Aprendizaje Supervisado

Mtro. René Rosado González
Director de Programa LTP

K-Vecinos Cercanos

K-Nearest Neighbors (K-NN)



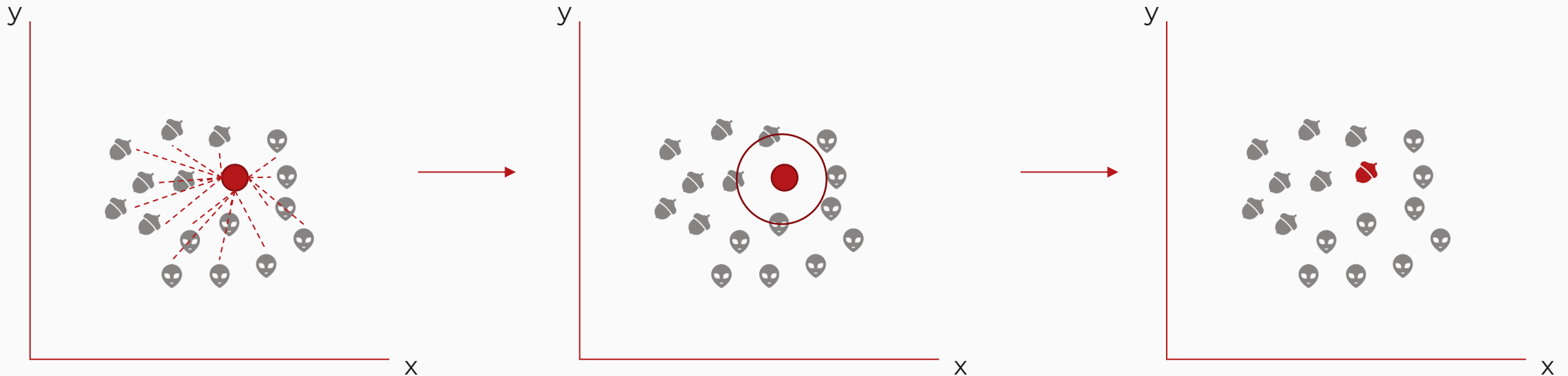
No hay suposiciones a priori sobre la relación entre las observaciones.

Todos los datos de entrenamiento son utilizados en la fase de prueba. Esto hace que el entrenamiento sea más rápido y la fase de prueba más lenta y costosa.

K-Vecinos Cercanos

K-Nearest Neighbors (K-NN)

1. Se calcula la distancia entre el nuevo punto y cada punto de entrenamiento.
2. Se seleccionan los k puntos de datos más cercanos (en función de la distancia).
3. El promedio de estos puntos de datos es la predicción final para el nuevo punto.



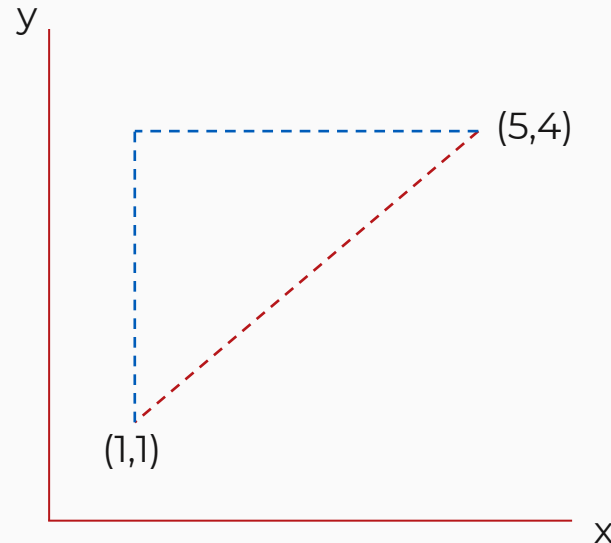
K-Vecinos Cercanos

Algunas medidas de distancia

Eculidiana

$$\sum_{i=1}^k \sqrt{(x_i - y_i)^2}$$

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



Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

$$|5 - 1| + |4 - 1| = 7$$

K-Vecinos Cercanos

Consideraciones

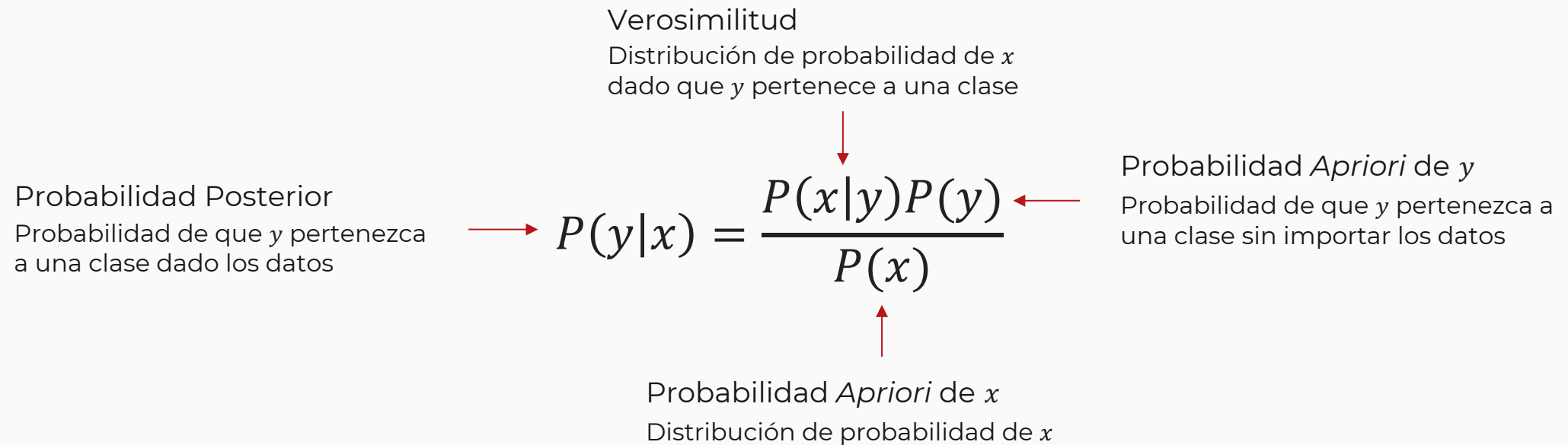
- Puede ser usado para regresión y clasificación.
- Funciona mejor con una cantidad menor de funciones que con una gran cantidad de funciones.
- El aumento de la dimensión también conduce al problema del sobreajuste.
- La investigación ha demostrado que en grandes dimensiones la distancia euclidiana ya no es útil.

Un ejemplo



Teorema de Bayes

Bayes Theorem

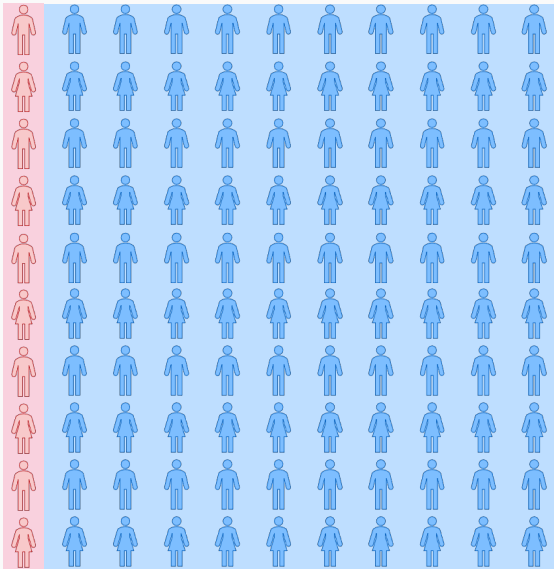


Teorema de Bayes

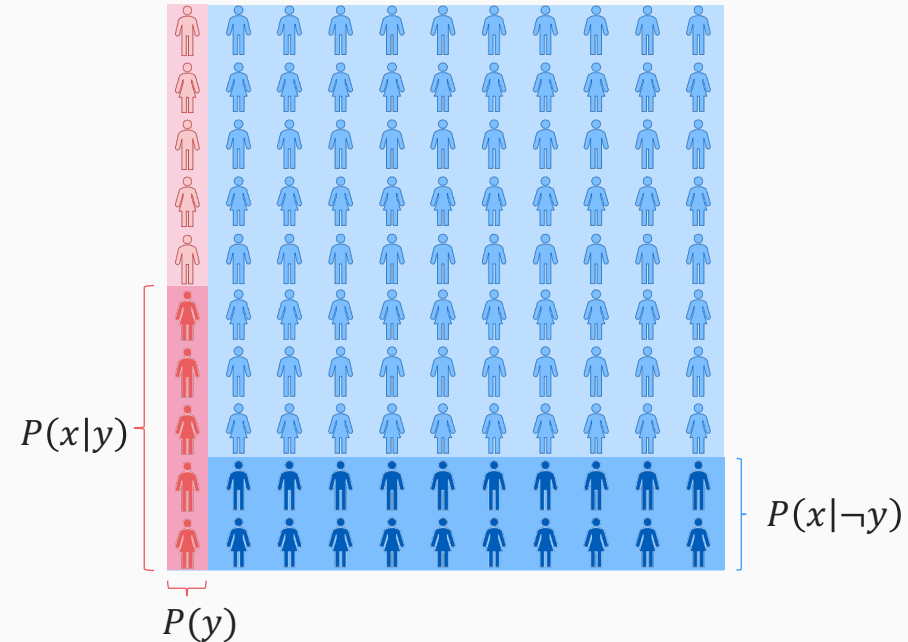
Bayes Theorem

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(x|y)P(y)}{P(x|y)P(y) + P(x|\neg y)P(\neg y)}$$

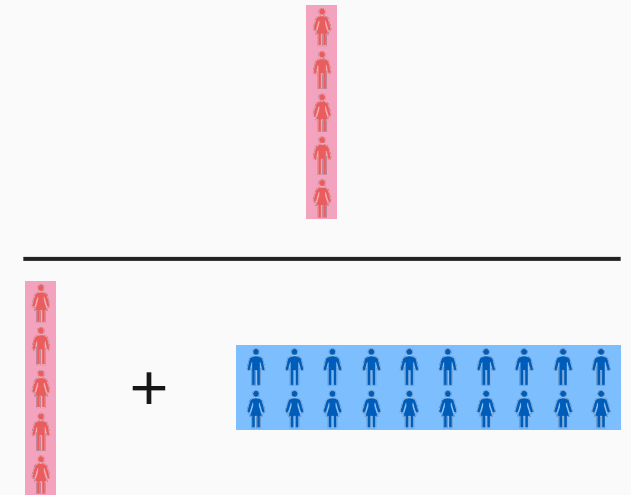
Creencia Inicial
(*apriori*)



Evidencia
(*versosimilitud*)



Creencia Actualizada
(*posterior*)



Bayes Ingenuo

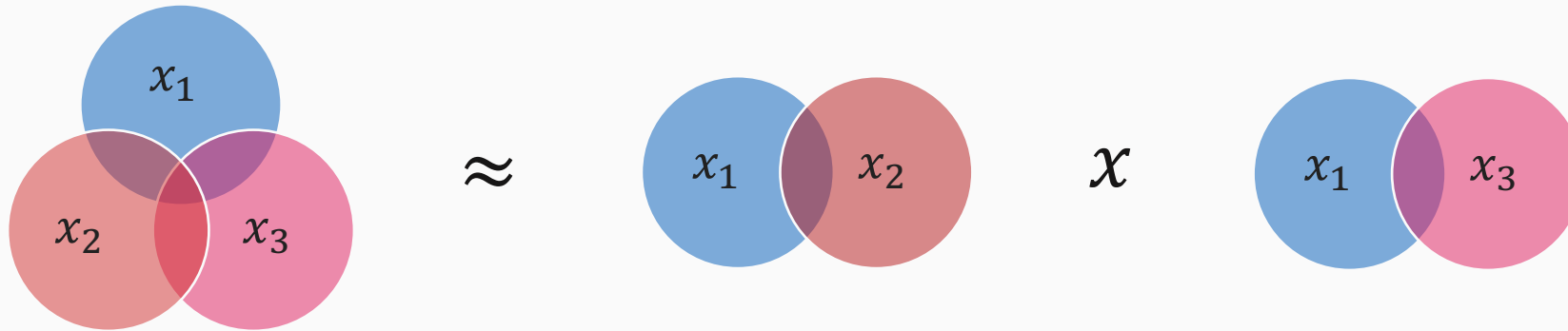
Naïve Bayes

- Es una técnica de clasificación estadística basada en el Teorema de Bayes.
- Asume que el efecto de una característica categórica particular en una clase es independiente de otras características categóricas.
- Algoritmo:
 1. Calcular la probabilidad inicial (*a priori*) de cada clase.
 2. Encontrar la probabilidad de verosimilitud con cada atributo para cada clase.
 3. Calcular la probabilidad posterior.
 4. Asignar la clase más probable.
- En el caso de variables continuas, podemos generar agrupaciones por particiones (bins)

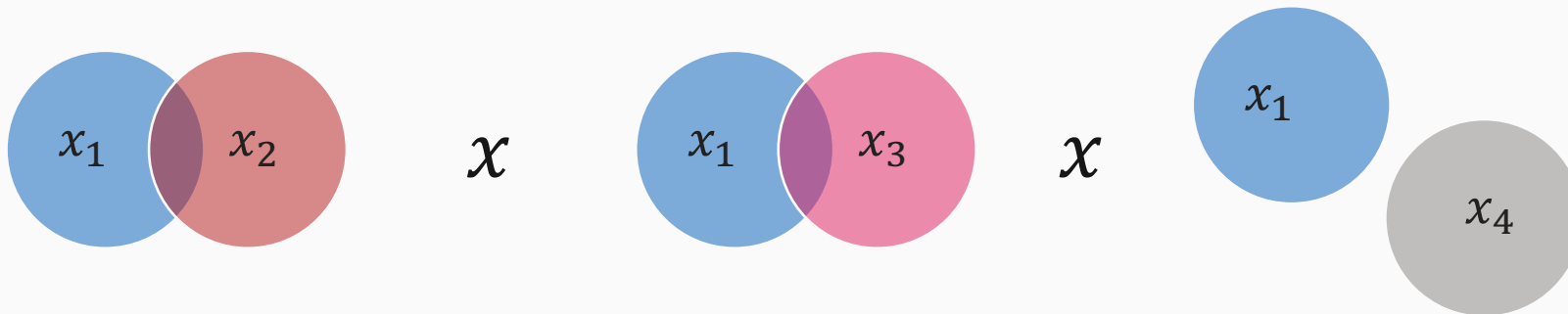
Bayes Ingenueo

Naïve Bayes

Múltiples Predictores



Problema con Observaciones Infrecuentes



Un ejemplo



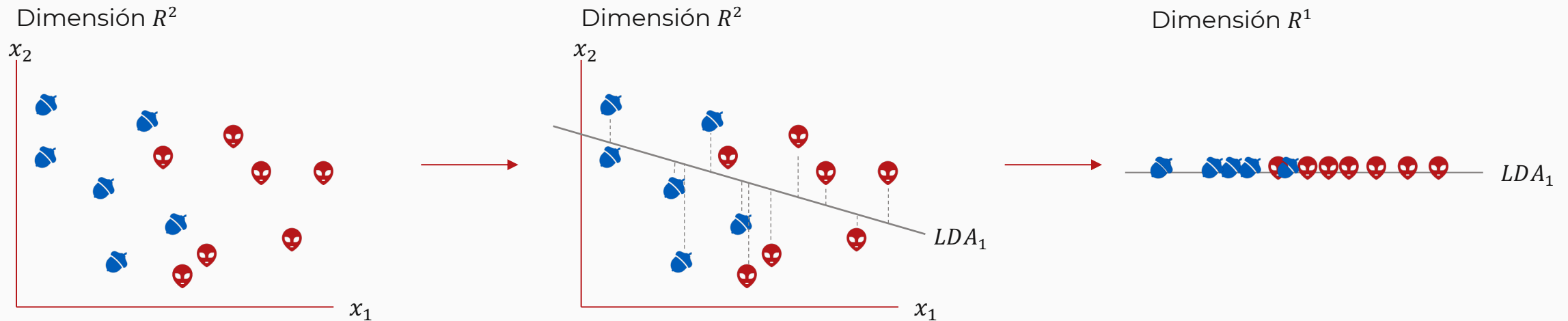
Analísis Lineal Discriminante

Linear Discriminant Analysis (LDA)

- Estima la probabilidad de que una observación, dado un conjunto de predictores, pertenezca a una clase.
- Permite la reducción de dimensionalidad en función de maximizar la capacidad de separar las clases considerando la máxima distancia entre las medias y la mínima dispersión entre las observaciones.
- Es una alternativa a la regresión logística cuando:
 - Las categorías a predecir están suficientemente separadas.
 - Se cuentan con pocas observaciones y los predictores siguen una distribución similar a la normal.

Anal is Linear Discriminante

Linear Discriminant Analysis (LDA)



Anal  s Lineal Discriminante

Linear Discriminant Analysis (LDA)

1. Calcular el vector de medias para todas las clases dentro de los datos

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j$$

2. Calcular la matriz de dispersi  n dentro de clase (*within-between-class*)

$$S_w = \frac{1}{N_i - 1} \sum_{j=1}^{n_i} (x_j - \mu_i)(x_j - \mu_i)^T$$

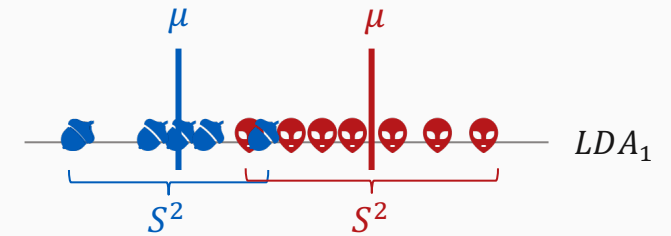
3. Calcular la matriz de dispersi  n entre clases (*in-between-class*)

$$S_b = \sum_{i=1}^c N_i (m_i - m)(m_i - m)^T$$

4. Calcular los vectores y valores propios de las matrices de dispersi  n.

5. Construir el subespacio transformado a partir de la matriz de valores propios tal que

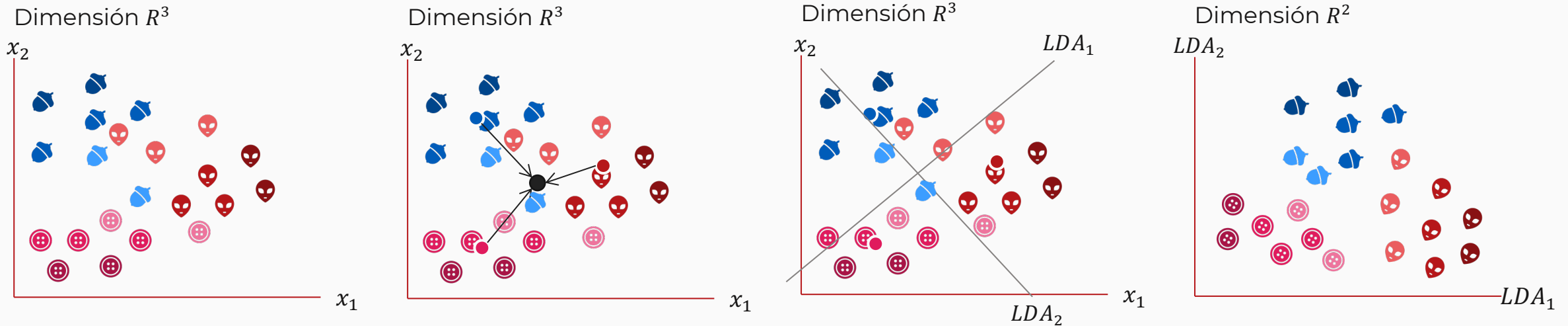
$$\lambda_{lda} = \operatorname{argmax}_p \frac{|\lambda^T S_b \lambda|}{|\lambda^T S_w \lambda|}$$



Minimizar $\longrightarrow \frac{(\mu - \mu)^2}{S^2 + S^2} \longleftarrow$ Maximizar

Anal  s Lineal Discriminante

Linear Discriminant Analysis (LDA)



Minimizar $\longrightarrow \frac{d^2 + d^2 + d^2}{S^2 + S^2 + S^2} \longleftarrow \text{Maximizar}$

Analísis Lineal Discriminante

Linear Discriminant Analysis (LDA)

Ventajas:

- Efectivo en para reducir dimensionalidad.
- Permite seleccionar variables en función de los pesos.
- Extensiones como el Análisis Cuadrático Discriminante permiten lidiar con problemas no lineales.
- Permite separar múltiples clases

Desventajas:

- Asume normalidad en los predictores.
- Generalmente no se desempeña bien cuando hay pocas categorías o poca varianza entre ellas.

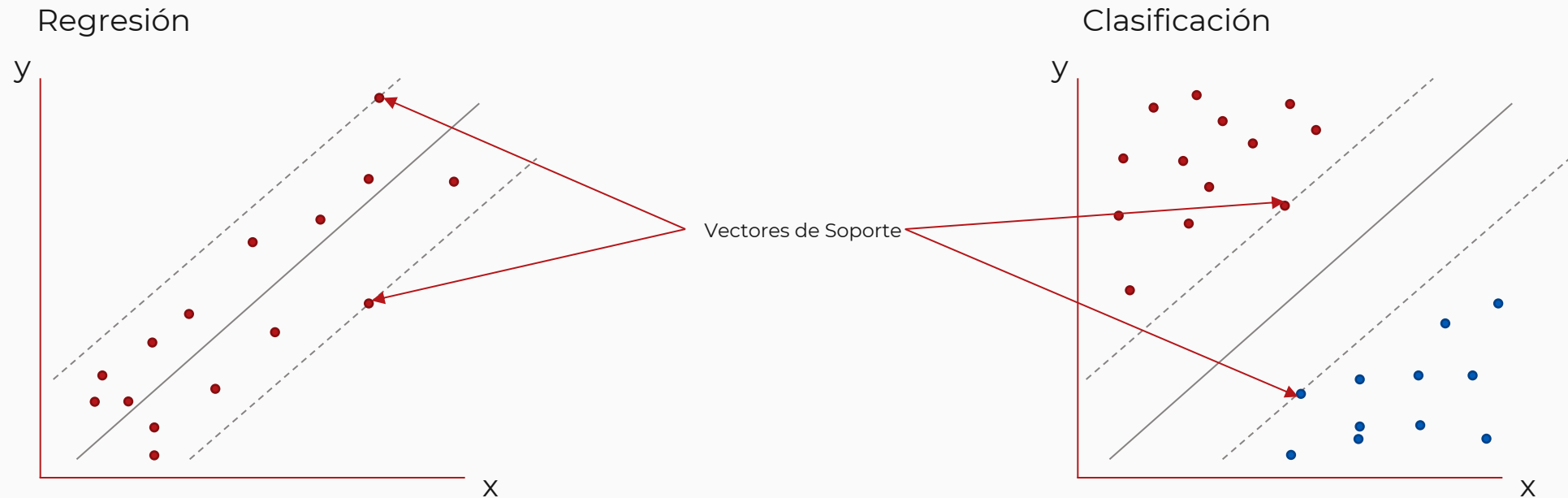
Un ejemplo



Máquinas de Soporte Vectorial

Support Vector Machines (SVM)

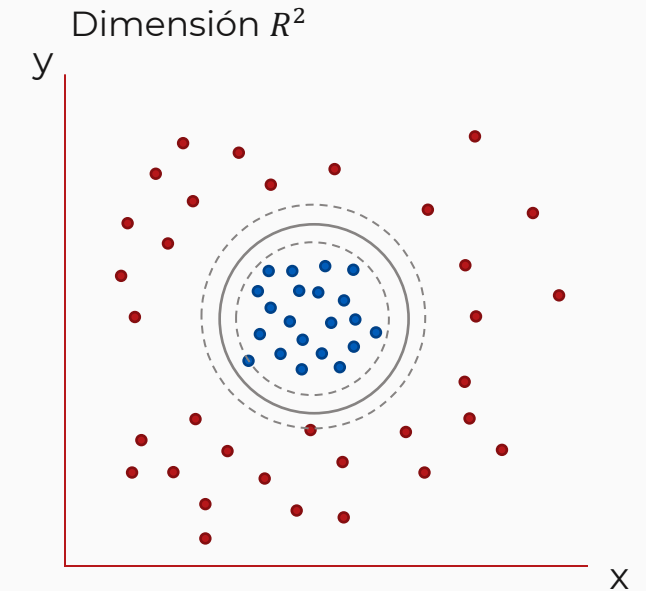
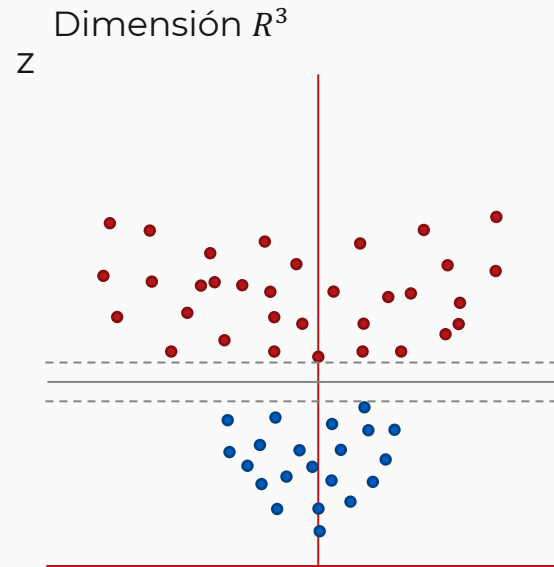
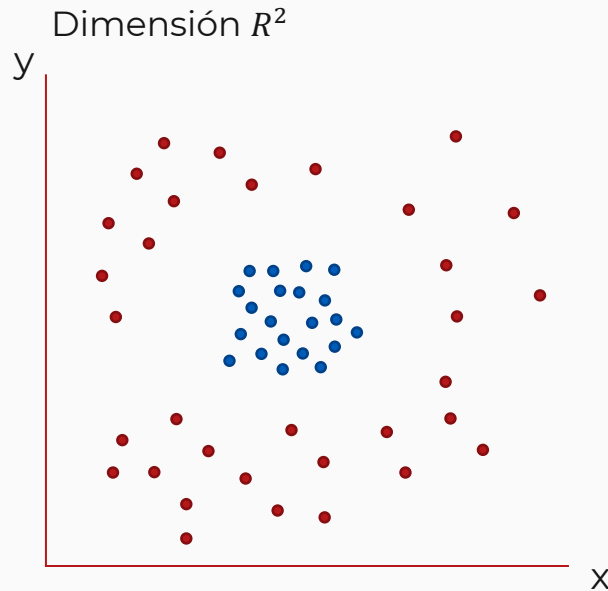
Se basan en la idea de ajustar la mayor cantidad de observaciones como sea factible dentro de un hiperplano y al mismo tiempo limitar las violaciones de un margen de decisión (maximal).



Máquinas de Soporte Vectorial

Support Vector Machines (SVM)

Dado que la máquina de soporte vectorial trabaja con kernels (funciones de núcleo) se desempeña bien ajustando problemas no lineales.



Máquinas de Soporte Vectorial

Support Vector Machines (SVM)

Ventajas:

- Efectivo en alta dimensionalidad.
- Eficiencia de memoria dado que solo se utiliza un subconjunto de los puntos de entrenamiento.
- Versatilidad en la separación no lineal de clases.
- Puede ser usado para regresión y clasificación.

Desventajas:

- En situaciones en las que el número de características para cada objeto excede el número de muestras de datos de entrenamiento, los SVM pueden funcionar mal.
- No existe una interpretación probabilística directa para la membresía del grupo.
- Por ahora solo existe clasificación binaria.

Un ejemplo

