

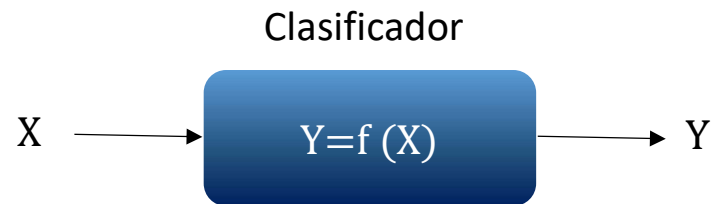
Aprendizaje Maquinal

Dr. Rubén Acevedo
ruben.acevedo@uner.edu.ar

Tecnicatura Universitaria
Procesamiento y Explotación de Datos

Clasificación

Definición



$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \\ c_j \end{pmatrix}$$

Instancia
(Dato, patrón)

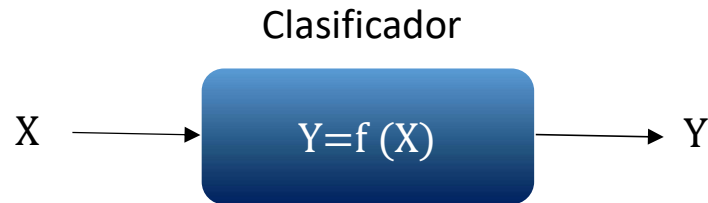
x_i se denominan *atributos*.
 c_j son las **clases**, con $j=1..K$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_M \end{pmatrix}$$

Salida

Clasificación

Definición



Se puede pensar un **clasificador** como una función $f: \mathbf{X} \rightarrow \mathbf{Y}$ que representa la correspondencia existente en una entrada \mathbf{X} (instancia) y una salida \mathbf{Y} .

\mathbf{Y} es nominal, es decir que puede tomar un conjunto de valores c_1, c_2, \dots, c_N denominados *clases*.

La *función* aprendida (modelo) será capaz de determinar la *clase* para cada nuevo ejemplo sin etiquetar, es decir dará un valor de \mathbf{Y} para cada valor de \mathbf{X} .

Clasificación

Características

Predictivos

Debe clasificar instancias previamente no etiquetadas. Un buen clasificador debe proporcionar predicciones precisas con un rápido tiempo de respuesta.

Descriptivos

Debe identificar las características que distinguen instancias de diferentes clases. Esto es particularmente útil para aplicaciones críticas, como el diagnóstico médico, donde es insuficiente tener un modelo que haga una predicción sin justificar cómo llega a tal decisión.

Clasificación

Características

Binarios

Los clasificadores binarios asignan cada instancia de datos a una de **dos** etiquetas posibles. Normalmente se denota como +1 y -1.

La clase positiva generalmente se refiere a la categoría de mayor interés en predecir correctamente en comparación con la clase negativa.

Multiclases

Si hay más de dos posibles etiquetas disponibles, entonces es un clasificador multiclases.

Algunos clasificadores fueron diseñados para clases binarias, en estos casos se pueden adaptar para tratar con problemas multiclases.

Clasificación

Características

Determinísticos

Un clasificador determinístico produce una etiqueta de valor discreto para cada dato o instancia que clasifica

Probabilísticos

Un clasificador probabilístico asigna un valor continuo entre 0 y 1 para indicar la probabilidad de que un dato o instancia pertenezca a un clase en particular, donde los puntajes de probabilidad para todas las clases suman 1.

Clasificación

Características

Lineales

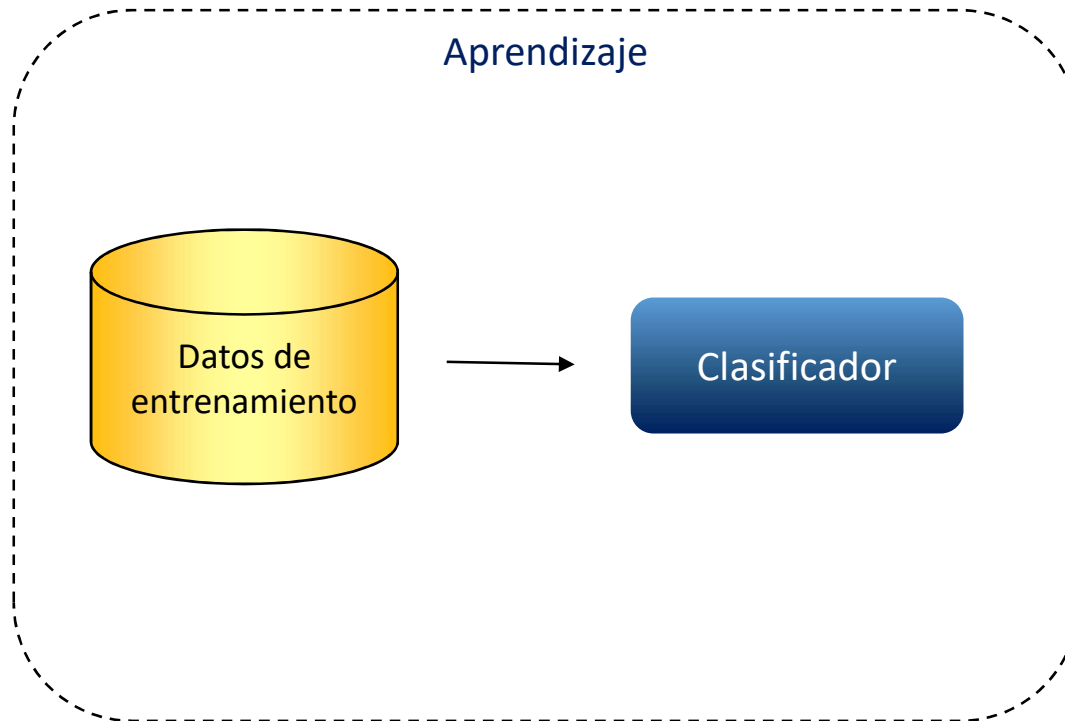
Un clasificador lineal utiliza un hiperplano de separación lineal para discriminar instancias de diferentes clases

No lineales

Un clasificador no lineal permite generar superficies de decisión no lineales más complejas

Clasificación

Aprendizaje



Clasificación

Aprendizaje

Tipos

Supervisado

Es el esquema más utilizado, este paradigma de aprendizaje está en función de salida deseada vs salida obtenida. Ejemplo: perceptron multicapa.

No supervisado

Se utilizan grandes cantidades de datos en los cuales se buscan regularidades estadísticas, definiendo en base a ellas categorías o clusters. Ejemplo: mapas de auto-organización.

Híbrido

Utiliza una combinación de los dos anteriores, bien en un orden o en el otro. Ejemplo. Redes de funciones de base radial

Clasificación

Aprendizaje

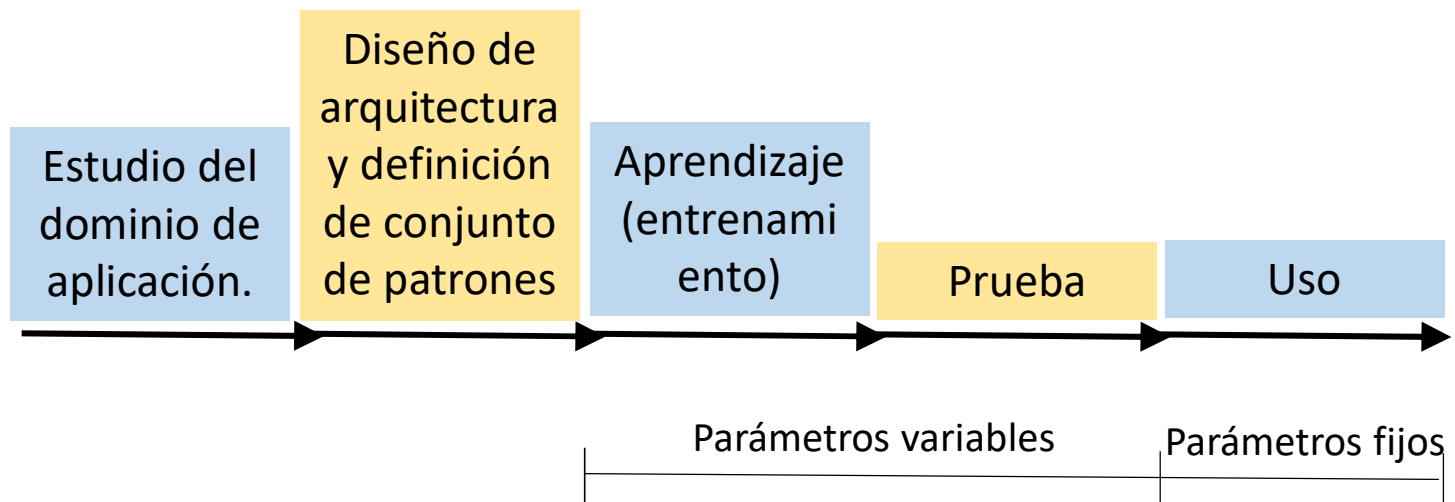
Tipos

On Line

En este caso el clasificador puede aprender durante su funcionamiento habitual.

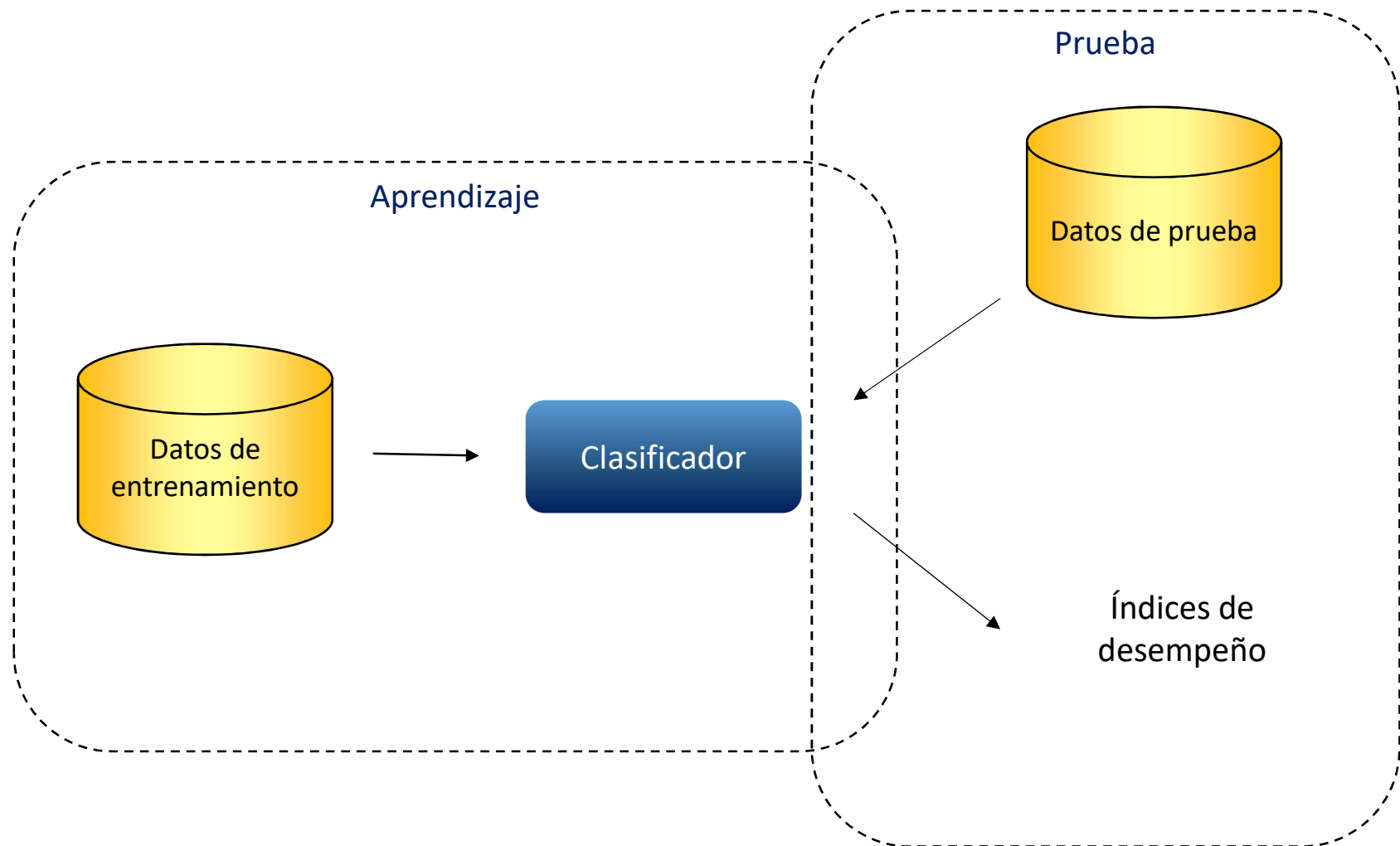
Off Line

Hay una fase de aprendizaje (*entrenamiento*) y una fase de uso (*operación*), existiendo un conjunto de datos de entrenamiento y un conjunto de datos de test o prueba.



Clasificación

Prueba



Clasificación

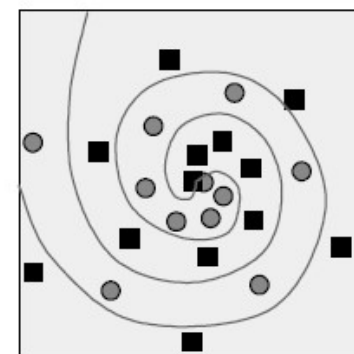
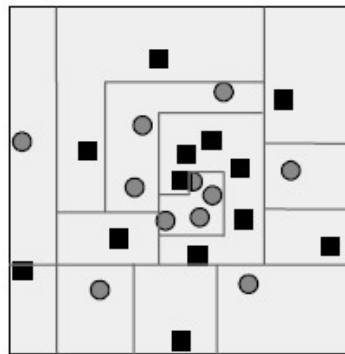
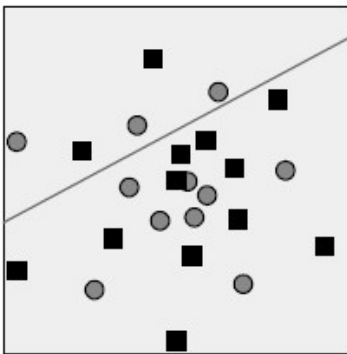
Prueba

Subajuste

El modelo (clasificador) no es capaz de ajustarse a los datos de entrenamiento de ninguna manera.

Sobreajuste

El modelo (clasificador) se ajusta perfectamente a los datos de entrenamiento pero no generaliza.



Clasificación

Prueba

En el proceso de selección de un modelo (clasificador) se utilizan criterios de evaluación del desempeño.

Algunas técnicas de validación

Hold-out

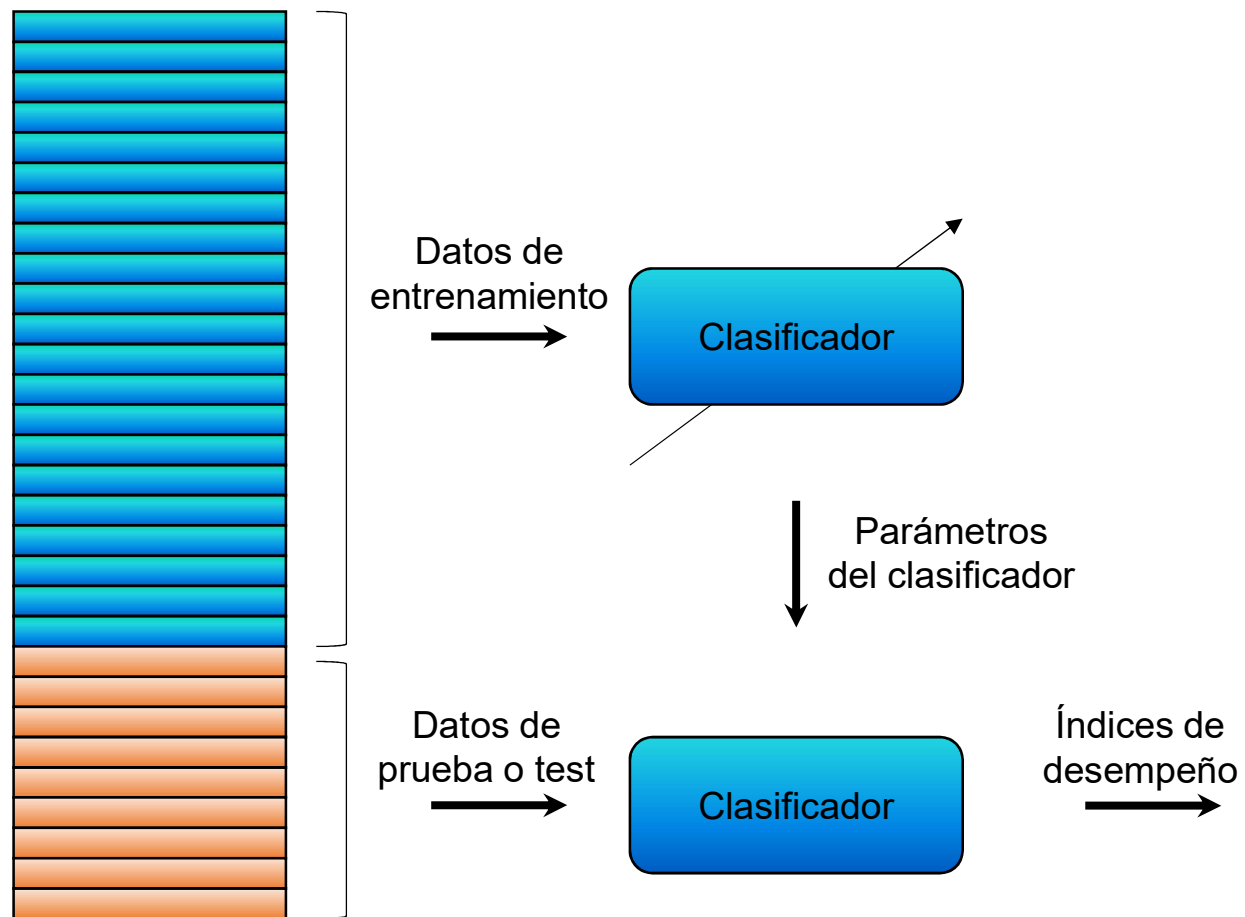
Validación cruzada

Leave-one-out

Clasificación

Prueba

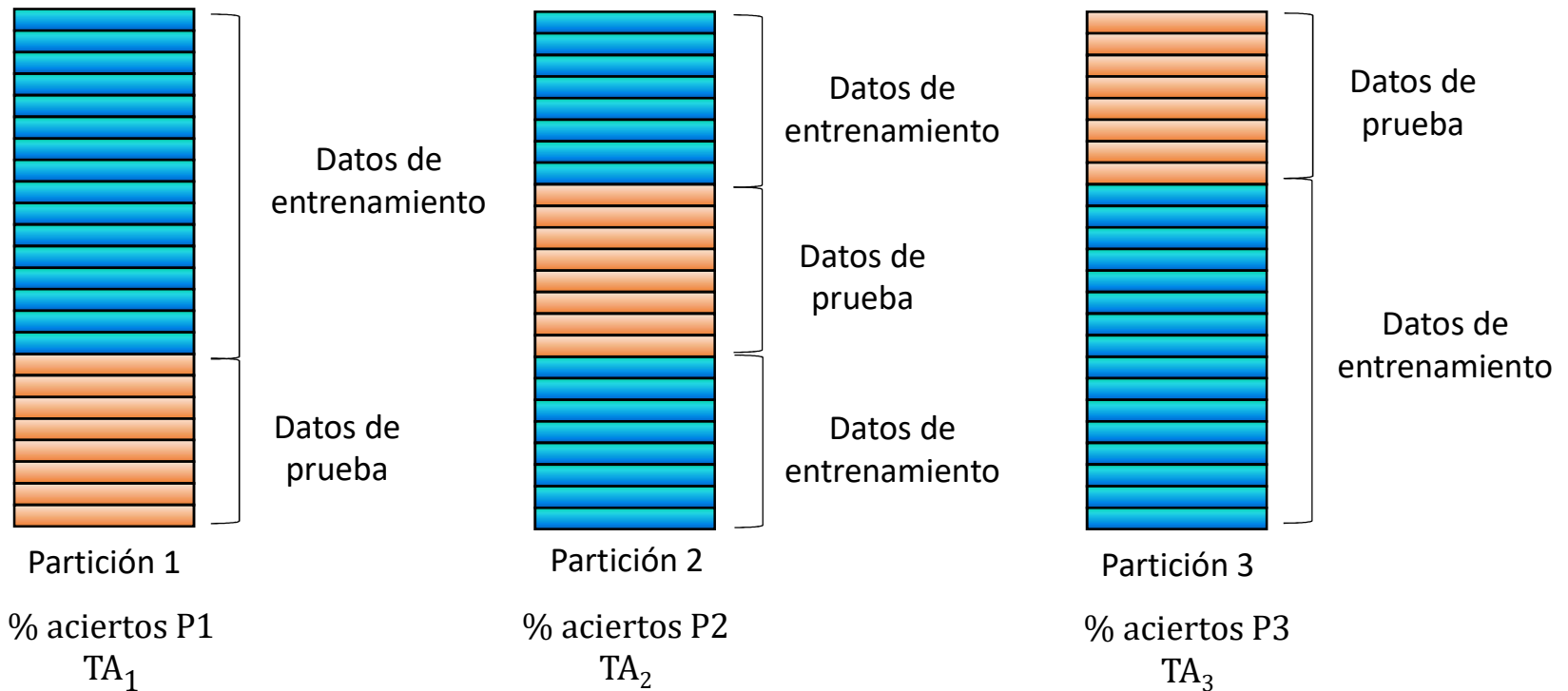
Hold out



Clasificación

Prueba

Validación cruzada



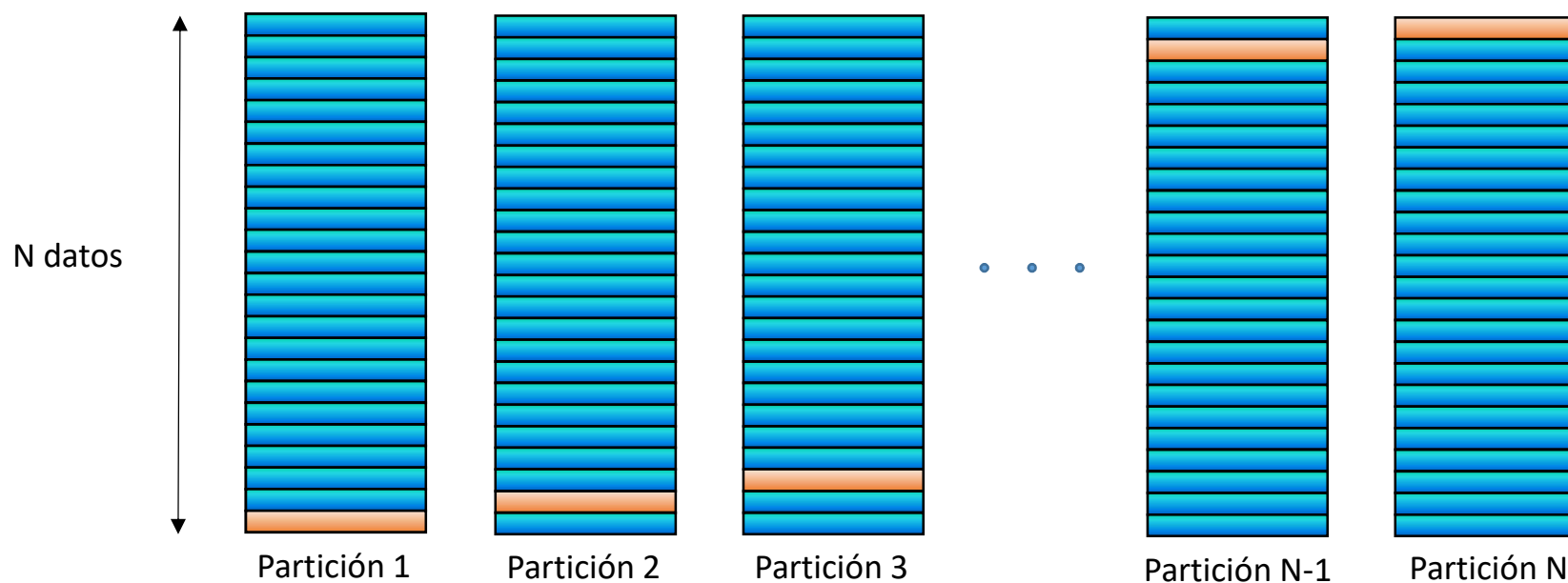
$$TA = \frac{TA_1 + TA_2 + TA_3}{3}$$

$$\text{Generalizando: } TA = \frac{\sum_{j=1}^K TA_j}{K}$$

Clasificación

Prueba

Leave one out



$$TA = \frac{\sum_{j=1}^N TA_j}{N}$$

Clasificadores basados en reglas

Prueba

Índices de desempeño

Matriz de confusión

Clasificador	Gold Standard		
		+	-
	+	VP	FP
	-	FN	VN

VP: verdaderos positivos

VN: verdaderos negativos

FP: falsos positivos

FN: falsos negativos

Clasificación

Prueba

Índices de desempeño

Accuracy (tasa de aciertos): porcentaje de predicciones correctas

$$Acc = \frac{VP + VN}{N}$$

Especificidad: porcentaje de aciertos de clasificados negativos

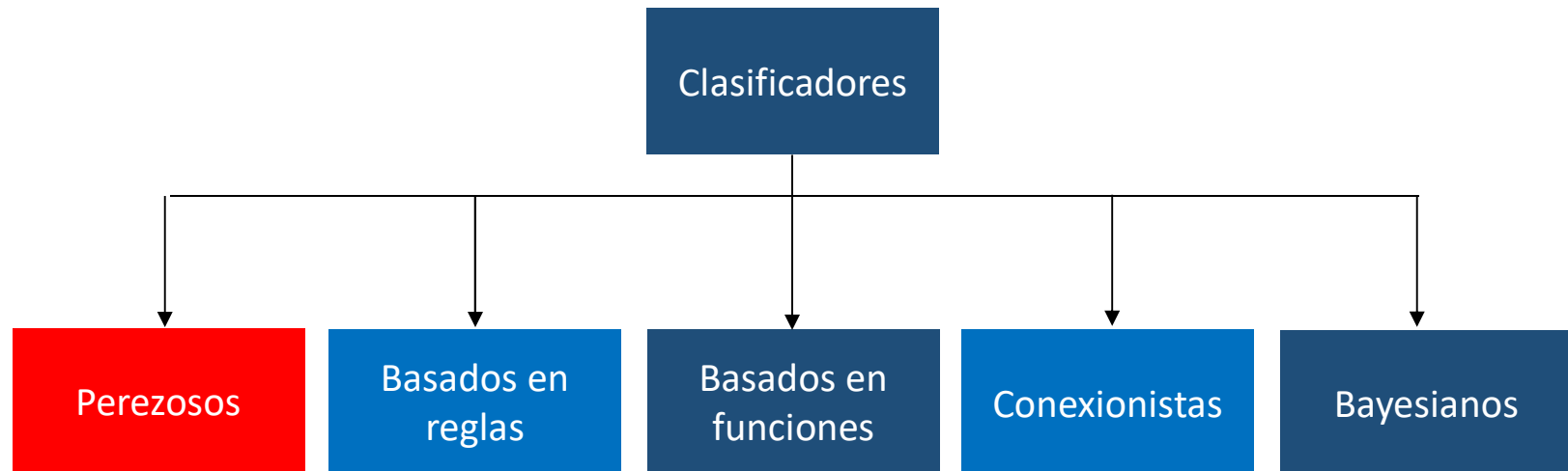
$$Especificidad = \frac{VN}{VN + FP}$$

Recall: porcentaje de aciertos de clasificados positivos

$$Sensibilidad = \frac{VP}{VP + FN}$$

Clasificación

Tipos



Clasificadores perezosos (*lazys*)

Definición

No **generan un modelo o función**, sino que realizan una optimización local.

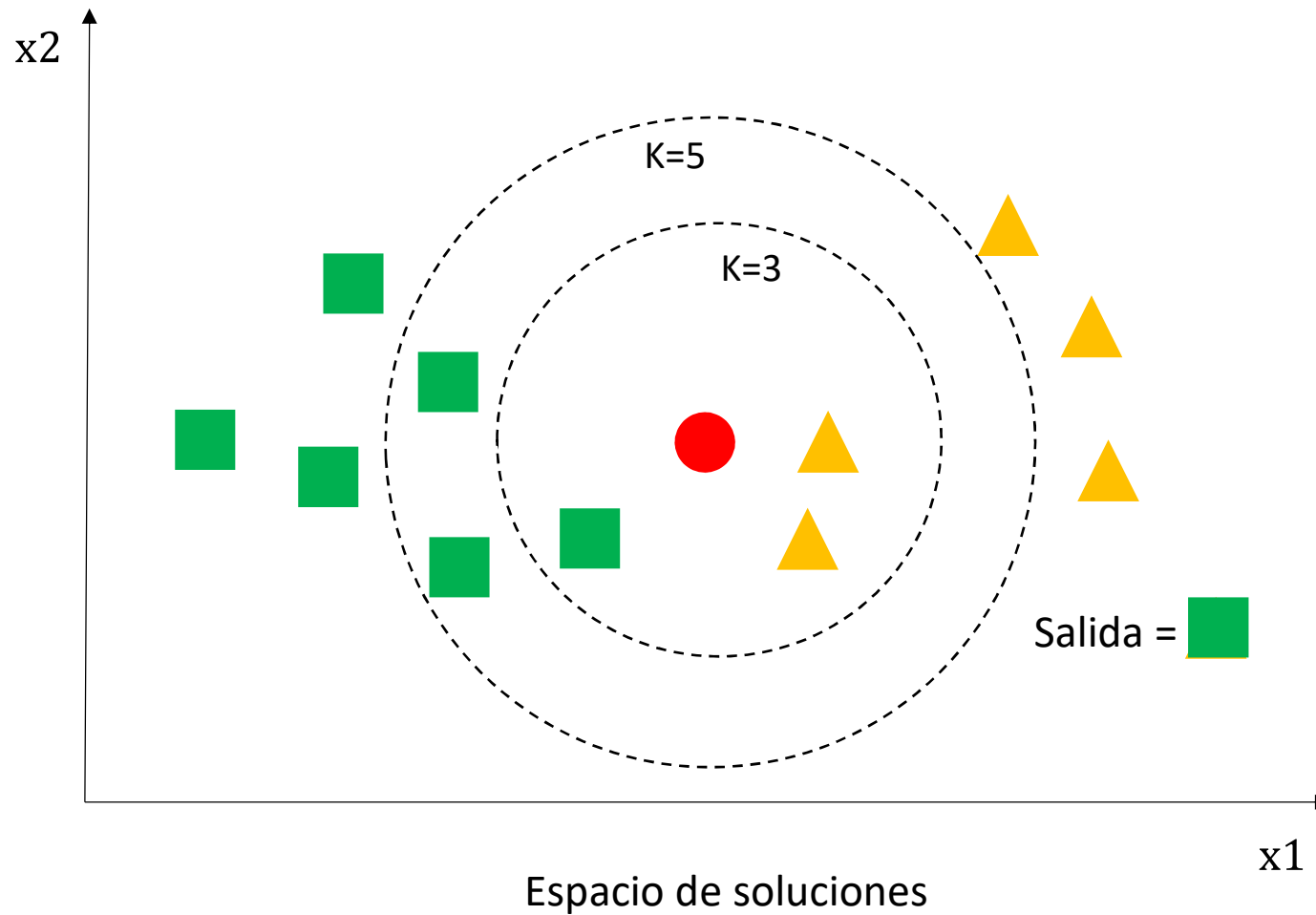
Son métodos que actúa para cada pregunta o predicción requerida.

Los ejemplos deben preservarse porque son necesarios para realizar cada predicción.

Una ventaja es que **no hay que entrenar** el modelo.

Clasificadores perezosos (*lazys*)

K vecinos mas cercanos

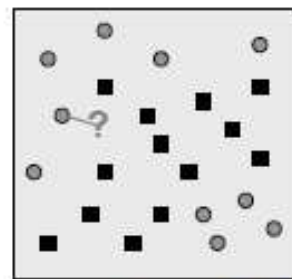


Clasificadores perezosos (*lazys*)

K vecinos mas cercanos

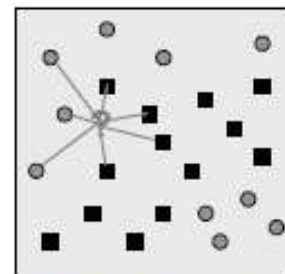
$X = (x_1, x_2, x_3, \dots, x_N)$, instancias de entrenamiento

$y \in C_k$, donde C_k son las clases del problema



1-nearest neighbor

Clasifica
círculo



7-nearest neighbor

Clasifica
cuadrado

Clasificación

K vecinos mas cercanos

Distancias

$$d(x, y) = \|x - y\| = \sum_{i=1}^k (x_i - y_i)^2, \text{ distancia Euclídea}$$

$$d(x, y) = \sum_{i=1}^k |x_i - y_i|, \text{ distancia Manhattan}$$

$$d(x, y) = \left[\sum_{i=1}^k (|x_i - y_i|)^4 \right]^{\frac{1}{4}}, \text{ distancia Minkowski}$$

Clasificación

K vecinos mas cercanos

Selección de K

