Ciencia y Tecnología
Secretaría de Ciencia, Humanidades, Tecnología e Innovación

CICESE

Laboratorio de Ciencia de Datos y Aprendizaje Automatico

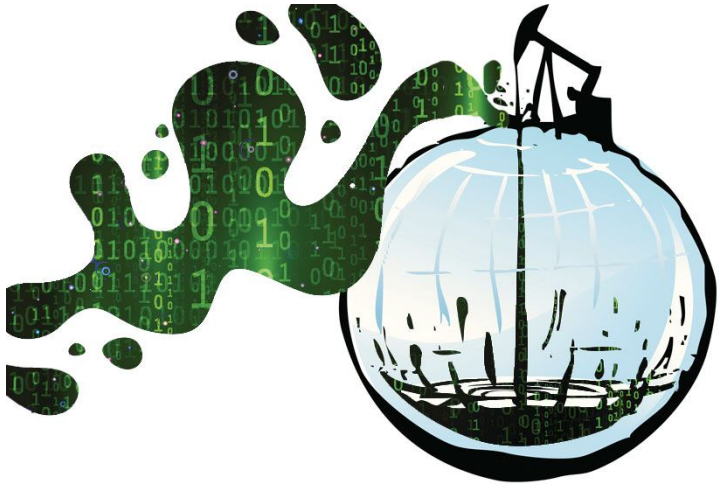
Curso IA desde Cero

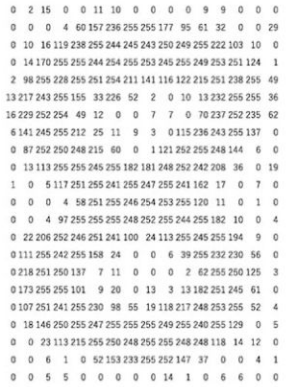
Dr. Irvin Hussein López Nava
M.C. Joan M. Raygoza Romero

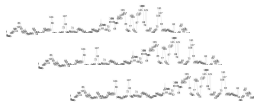
Departamento de Ciencias de la Computación
Centro de Investigación Científica y de Educación Superior de Ensenada
Edición 2025

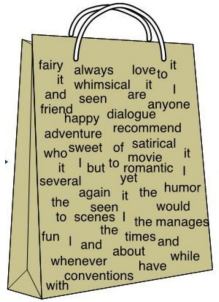
Recursos

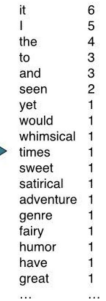
Datos



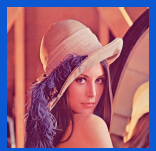




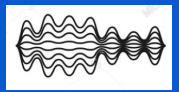




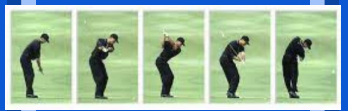
Imágenes



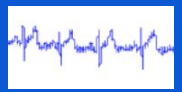
Audio



Video

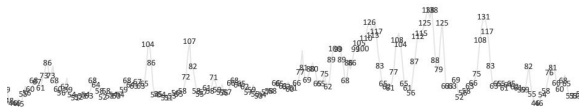


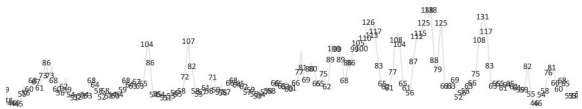
Señales



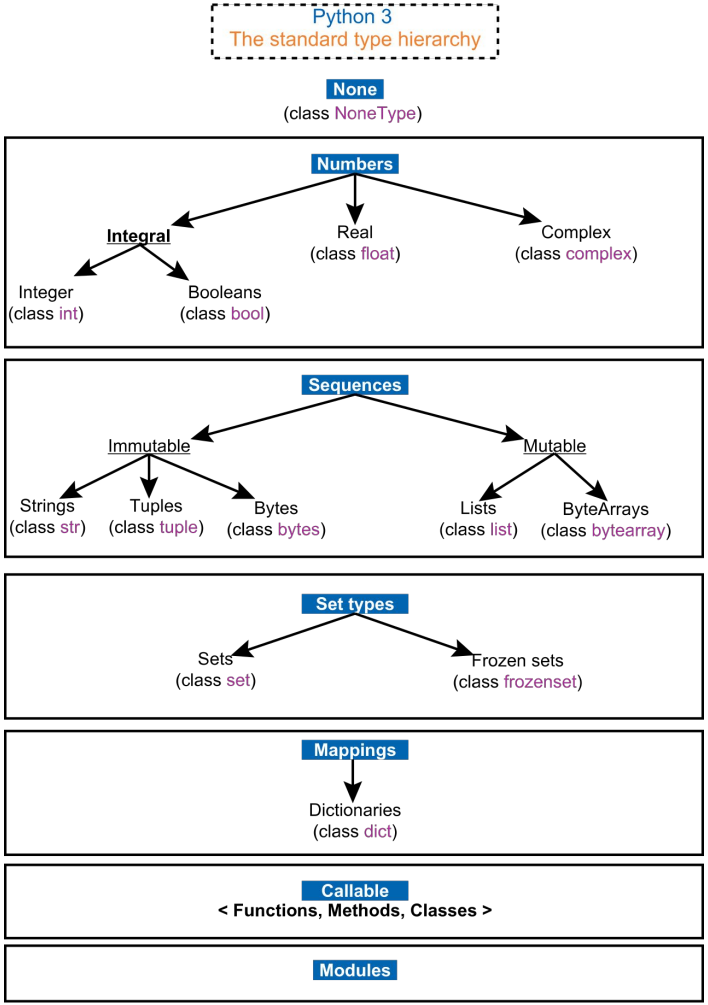
Texto

Lorem ipsum dolor sit dolor amet, consectetur nec adipiscing elit, sed do ipsum eiusmod tempor.

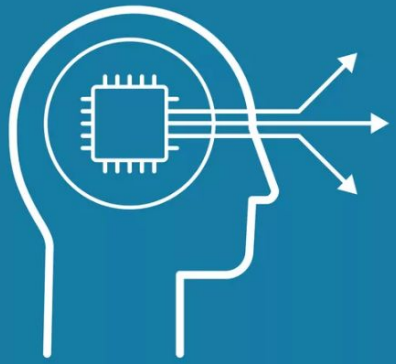




Datos en Python

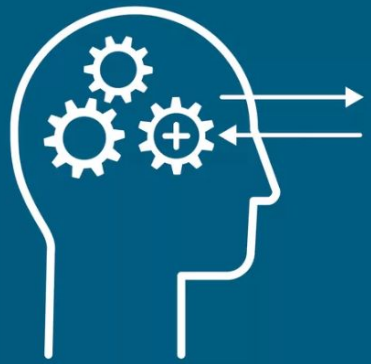


1950 1980 2010



ARTIFICIAL INTELLIGENCE

ENGINEERING OF MACHINES
THAT MIMIC COGNITIVE FUNCTIONS



MACHINE LEARNING

ABILITY TO PERFORM TASKS
WITHOUT EXPLICIT INSTRUCTIONS
AND RELYING ON PATTERNS

DEEP LEARNING

MACHINE LEARNING BASED
ON ARTIFICIAL NEURAL NETWORKS

¿Qué nos interesa más ahora?

Aprendizaje automático

Obj:

Modelos predictivos

Reconocimiento de patrones/estructuras

Clasificación natural

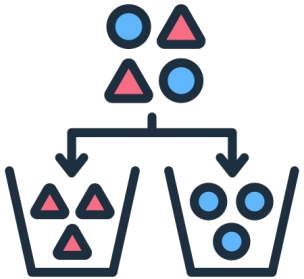
Un primer caso

Modelo de clasificación

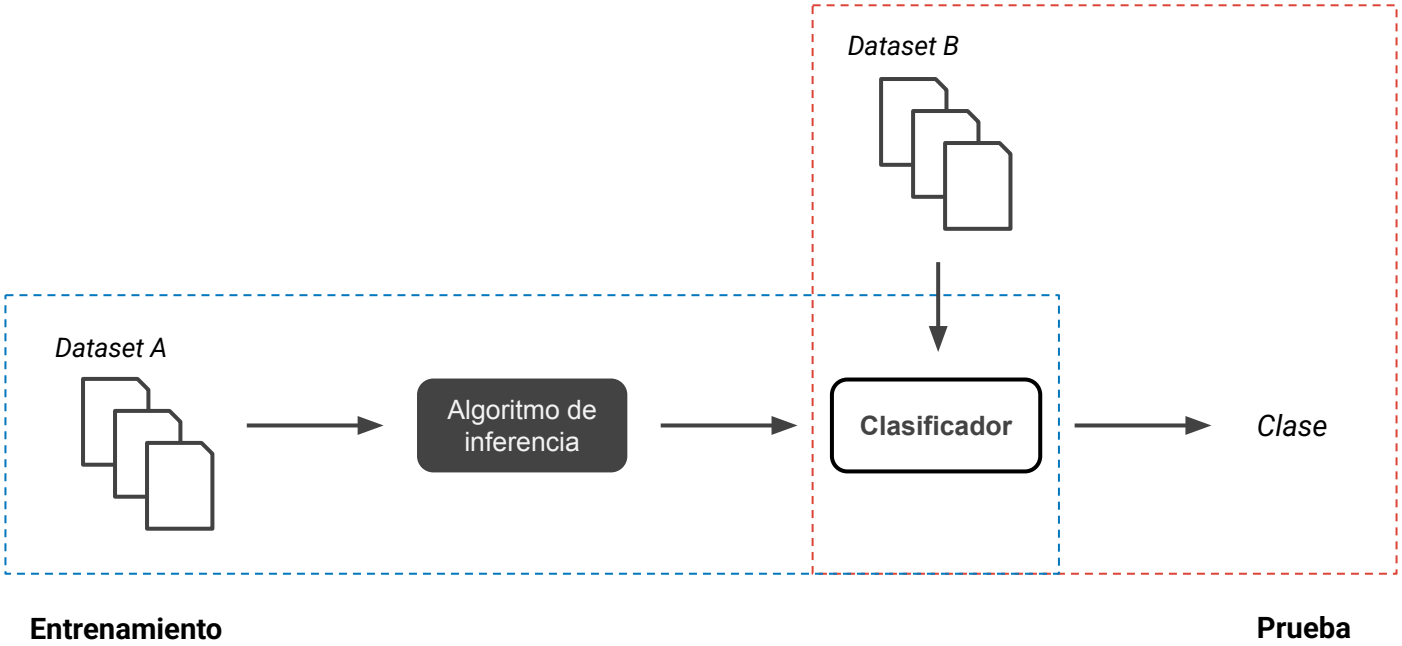
gato

¿Qué es clasificación?

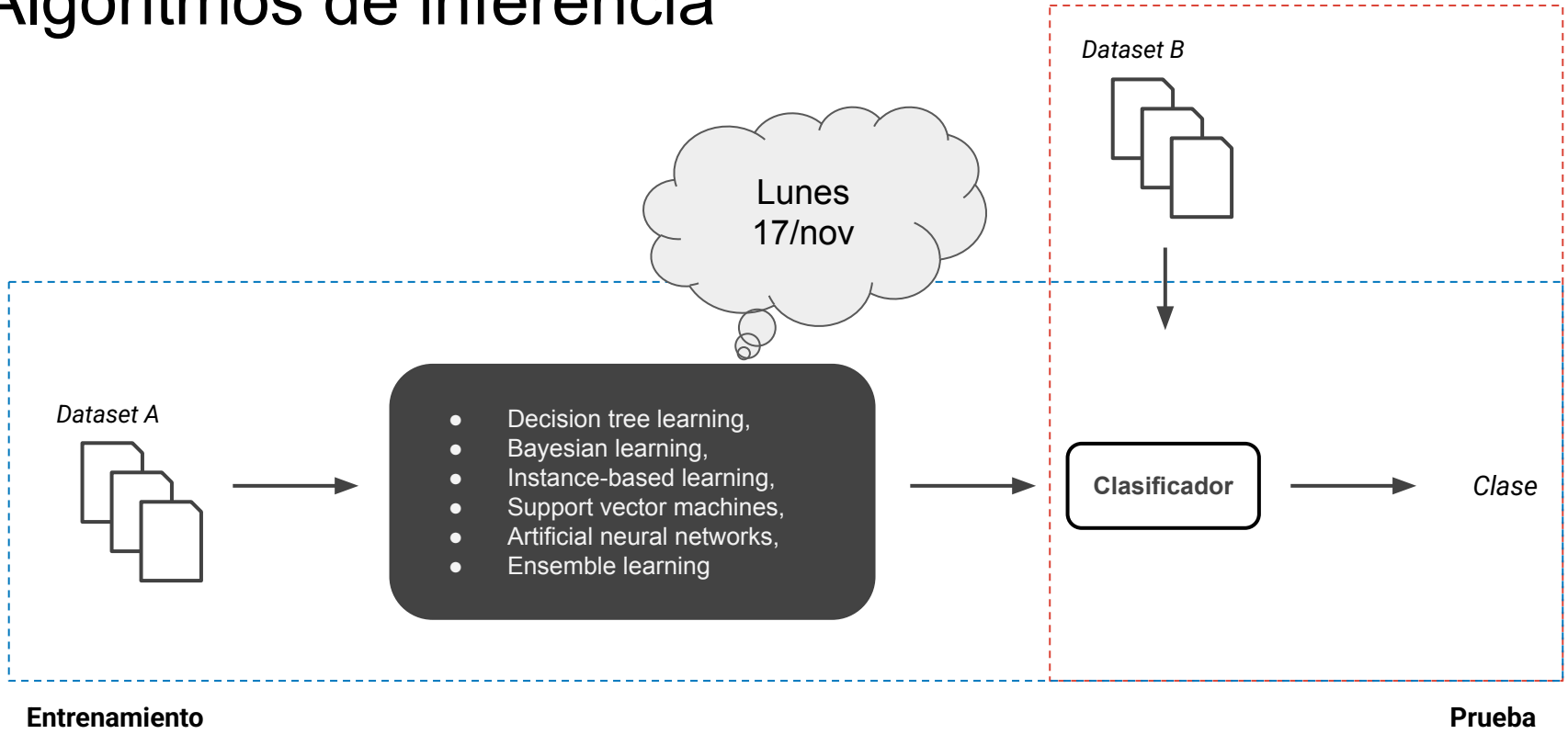
- En el **aprendizaje supervisado**, la **clasificación** es el proceso de identificar a cuál de un conjunto de clases pertenece una **nueva observación** (fase de prueba), a partir de un conjunto de **datos de entrenamiento** que contiene observaciones cuya clase se conoce (fase de entrenamiento).



Clasificador general



Algoritmos de inferencia

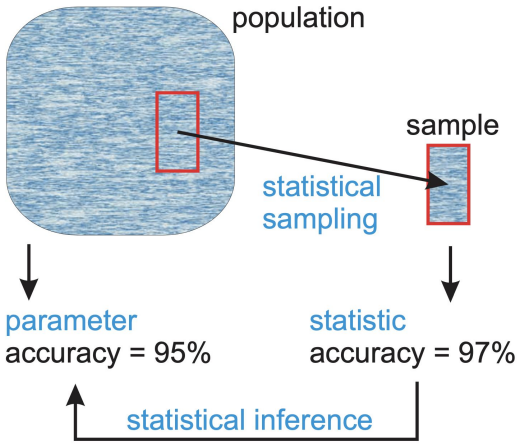


¿Hasta qué punto confiamos en el clasificador aprendido?

- Los clasificadores (tanto supervisados como no supervisados) se aprenden (entrenan) en un conjunto de **datos de entrenamiento** **finito**.
- Un **clasificador**, o modelo, aprendido debe probarse experimentalmente en una prueba diferente (datos de prueba).
- El rendimiento experimental sobre los **datos de prueba** es una aproximación al rendimiento en datos desconocidos – comprueba la capacidad de generalización del clasificador.
- Se necesita una función que evalúe experimentalmente el rendimiento del clasificador, e.g., su tasa de error, precisión, sensibilidad, especificidad.
    - Es necesario comparar los clasificadores experimentalmente.

La evaluación es una prueba de hipótesis

- La evaluación debe tratarse como una **prueba de hipótesis** en estadística.
- El valor del parámetro poblacional debe deducirse estadísticamente a partir de las estadísticas de la muestra (i.e., un conjunto de entrenamiento).



Partición de los datos

Uso de los datos: conjuntos de entrenamiento y de prueba

Problema: sólo se dispone de datos finitos y deben utilizarse tanto para el entrenamiento como para las pruebas.

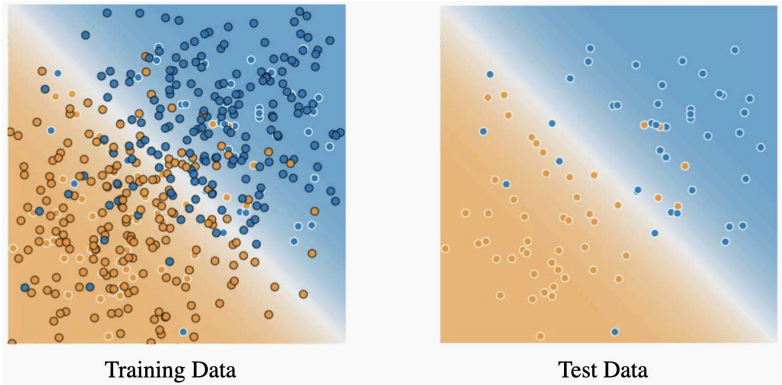
- Un mayor número de datos de entrenamiento mejora la generalización.
- Más datos de prueba dan una mejor estimación de la probabilidad de error de clasificación.
- **NUNCA** se debe evaluar el rendimiento de los clasificadores en función de los **datos de entrenamiento**: la conclusión tendría un sesgo optimista.

División de los datos

Partición (división) del conjunto finito de datos disponible en subconjuntos de entrenamiento/prueba:

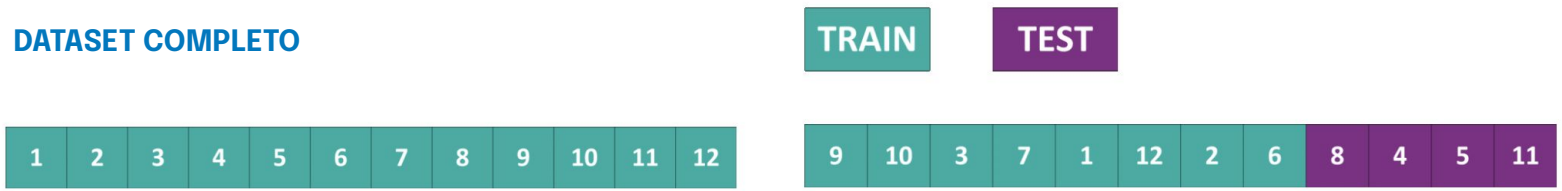
- Hold out.
- Cross validation.
- Bootstrap.

Una vez finalizada la evaluación, todos los datos disponibles pueden utilizarse para entrenar el clasificador final.



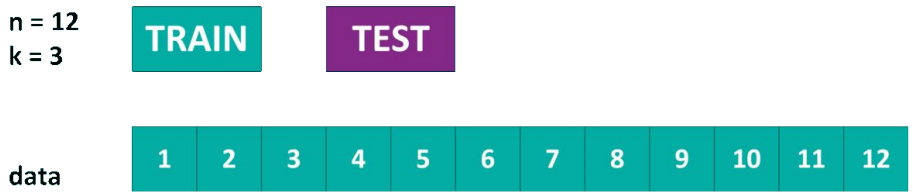
Hold out method

- Los datos se dividen **aleatoriamente** en dos conjuntos independientes.
- El **conjunto de entrenamiento** (e.g., 2/3 de los datos) para la construcción del modelo estadístico, i.e., el aprendizaje del clasificador.
  - Generalmente la proporción de los datos es mayor al conjunto de prueba.
- El **conjunto de prueba** (e.g., 1/3 de los datos) se utiliza para estimar la precisión del clasificador.



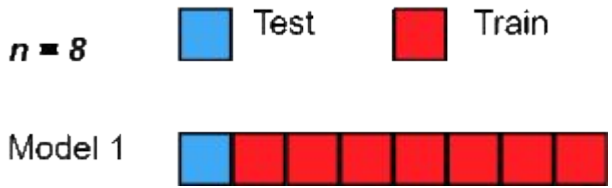
K-fold cross validation

- El **conjunto de entrenamiento** se divide **aleatoriamente** en *k* conjuntos disjuntos de igual tamaño en los que cada parte tiene aproximadamente la misma distribución de clases.
- El clasificador se evalúa *k* veces, cada vez con un conjunto diferente que se utiliza como **conjunto de prueba**.
- El rendimiento del clasificador es la media de estos *k* conjuntos.



Leave-one-out

- Un caso especial de validación cruzada *k*–fold con *k* = *n*, donde *n* es el número total de muestras del conjunto de datos.
  - Se realizan *n* experimentos utilizando n - 1 muestras para el entrenamiento y la muestra restante para las pruebas.
- Es bastante costosa desde el punto de vista computacional.



Métricas de evaluación

Retomando el *accuracy*

- La **exactitud** es el porcentaje de clasificaciones correctas.
- La **tasa de error** es el porcentaje de clasificaciones incorrectas.
- *Accuracy = 1 – Error rate*.

Problemas con estas métricas:

- Asume costes iguales para la clasificación errónea.
- Supone una distribución de clases relativamente uniforme (e.g., 0.5% de pacientes con una determinada enfermedad).

Se pueden derivar otras métricas de la matriz de confusión.

Matriz de confusión (tipos de errores)

- Una matriz de confusión, o matriz de error, es un diseño de tabla específico que permite visualizar el rendimiento de un modelo de clasificación.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN (error type I)
	Negative	FP (error type II)	TN

El primer tipo de error es **el rechazo** de una hipótesis nula verdadera.
El segundo es **el NO rechazo** de una hipótesis nula falsa.

Como ejemplo, un error de tipo I corresponde a condenar a un acusado inocente; mientras que un tipo II corresponde a absolver a un criminal.

Matriz de confusión (caso binario)

		Predicted		
		Positive	Negative	
Actual	Positive	True positive (TP)	False negative (FN)	Sensitivity
	Negative	False positive (FP)	True negative (TN)	Specificity
		Precision		F-measure

- Sensitivity, recall, o true positive rate (TPR): $$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$
- Specificity o true negative rate (TNR): $$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$
- Precision o positive predictive value (PPV): $$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$
- Accuracy (ACC): $$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$
- F1 score es la media armónica entre precision and sensitivity: $$\mathrm{F}_1 = 2 \cdot \frac{\mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

Costes desiguales de las decisiones

- **Diagnóstico médico:** El coste de una indicación falsa de cáncer de mama en el cribado de la población es menor que el coste de pasar por alto una enfermedad verdadera.
- **Defensa contra misiles aéreos:** El coste de no detectar un ataque real es mucho mayor que el de una falsa alarma.

Problema de la distribución de clases desconocida

En muchas circunstancias, se desconoce la distribución de clases, e.g., un filtro de spam de correo electrónico. Los modelos estadísticos deben aprenderse de antemano.

- Diagnóstico médico: 95 % sano, 5 % enfermedad.
- Comercio electrónico: el 99 % no compra, el 1 % compra.
- Seguridad: el 99.999 % de los ciudadanos no son terroristas.

Una situación similar ocurre con los clasificadores multiclase. La clase mayoritaria puede acertar el 99 %, pero es inútil.

Dealing with the imbalance: sampling

Construir un conjunto de entrenamiento equilibrado para entrenar el clasificador.

- Seleccione aleatoriamente el número deseado de casos de clase minoritaria.
- Añada el mismo número de casos de clase mayoritaria seleccionados aleatoriamente.

Construya un conjunto de prueba equilibrado (diferente del conjunto de entrenamiento) para probar el clasificador.

Dealing with the imbalance: augmentation

Construir un conjunto de entrenamiento equilibrado para entrenar el clasificador.

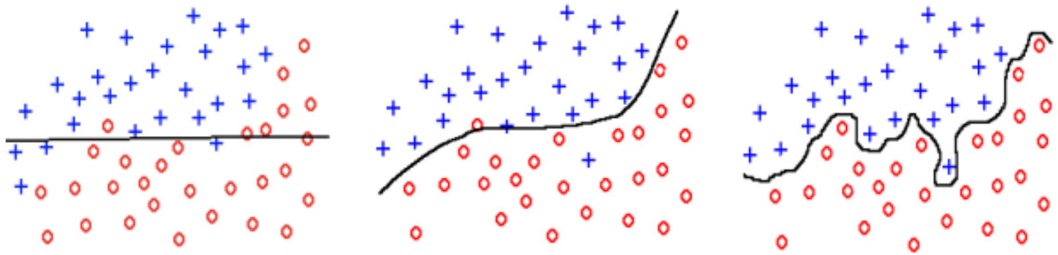
- Seleccione el número deseado de instancias de clase mayoritaria.
- Añada el mismo número de casos sintéticos de clases minoritarias.

Construya un conjunto de prueba equilibrado (**sin incluir datos sintéticos**) para probar el clasificador.

¿Es suficiente una métrica para evaluar el rendimiento de un clasificador?

Riesgo de sobreajuste

- **Aprender** de los datos de entrenamiento con demasiada precisión suele dar lugar a malos resultados de clasificación en nuevos datos.
- ¡El clasificador debe tener la capacidad de **generalizar**!



¿Preguntas?

hussein@cicese.mx