

Padrón: _____ Nombre y Apellido: _____ Grupo : _____

Parcial 20-11-2025

Ejercicio 1

Explique la diferencia entre variables cuantitativas y variables cualitativas. ¿Cómo se relacionan con los distintos tipos de problemas en machine learning (clasificación, regresión y agrupamiento)? Incluya ejemplos para cada caso.

Diferencia entre variables cuantitativas y cualitativas

- **Cuantitativas:** son numéricas y permiten operaciones matemáticas. Representan cantidades. Ejemplos: precio, edad, ingresos, temperatura.
- **Cualitativas:** describen categorías o cualidades, no son numéricas. Ejemplos: color, estado civil, tipo de producto, país.

Relación con tipos de problemas de ML

- **Clasificación:** objetivo cualitativo (categorías). Ej: predecir si un cliente comprará (sí/no).
- **Regresión:** objetivo cuantitativo (valor numérico). Ej: predecir el precio de una casa.
- **Agrupamiento:** se usan ambas según el algoritmo, se agrupan por similitud. Ej: segmentar clientes por edad (cuant.), país (cualit.) y gasto (cuant.)

Ejercicio 2

Al analizar un dataset de ventas online, se observa que el histograma de precios de los productos presenta una distribución fuertemente asimétrica con sesgo positivo. ¿Qué transformación de la variable podría aplicarse antes de entrenar un modelo para reducir dicha asimetría?

La variable precio tiene distribución asimétrica con sesgo positivo (cola hacia la derecha).

Transformación recomendada: **Log-transform (logaritmo natural)**

Reduce la asimetría y estabiliza la varianza.

Ejercicio 3

Se entrenó un modelo de clasificación (MODELO A) para detectar personas enfermas (clase 1) y sanas (clase 0). Luego se evaluó el modelo en los datos de test y se obtuvo la siguiente matriz de confusión:

True	0	10	8
	1	9	17
		0	1
		Predicted	

Padrón: _____ Nombre y Apellido: _____ Grupo : _____

Luego se entrenó un segundo modelo de clasificación (MODELO B), **se evaluó sobre los mismos datos de test** y se obtuvieron los siguientes resultados: **True Positive: 8 False Positive: 9**
¿Qué modelo (A o B) elegiría en términos de Accuracy? ¿Cuál en términos de *recall*?

Modelo A

Accuracy = 27/44 Recall = 17/26

Modelo B

TP: 8 FP: 9 TN: 9 FN: 18

Accuracy = 17 / 44 Recall = 8 / 26

El modelo A tiene mejores métricas.

Ejercicio 4

Mencione y grafique dos funciones de activación comúnmente utilizadas en redes neuronales.
Mencione y explique dos técnicas utilizadas en redes neuronales para evitar el *overfitting*.

Funciones de activación

- ReLU : $f(x) = \max(0, x)$
- Sigmoidea : $f(x) = 1 / (1 + e^{(-x)})$

Dos técnicas para evitar overfitting

- Regularización (L1/L2): penaliza pesos grandes genera modelos más simples.
- Dropout: apaga aleatoriamente neuronas durante el entrenamiento evita dependencia excesiva.

Otras válidas: early stopping, data augmentation, etc

Ejercicio 5

¿Qué técnica de reducción de la dimensionalidad utilizaría si quiere preservar los *clusters*? ¿Cuál utilizaría si sospecha que los datos se agrupan en una **variedad** del espacio?

Preservar cluster: TSNE

Variedad en el espacio: ISOMAP

Ejercicio 6

En relación con el dataset de **estaciones meteorológicas de Australia** utilizado en el Ejercicio 2, indique qué características podrían afectar las métricas obtenidas por los modelos. ¿Qué medidas podrían implementarse para mitigar este problema?

1 - Desbalance de clases / Técnicas de sampling

2 - Valores Faltantes / Técnicas de imputación y/o selección de variables