

1. ¿Cuál es la diferencia entre métodos de aprendizaje supervisado y no supervisado?

Supervisado: Son datos etiquetados, el modelo aprende a partir de un conjunto de datos donde se sabe cuál es su categoría.

No Supervisado: Son datos no etiquetados, se deben encontrar patrones o estructuras para poder agruparlos.

2. Observa los siguientes dataset dummies:

- DATASET 1

ID	EEG_A (0.5–4Hz)	EEG_θ (4–8Hz)	EEG_α (8–12Hz)	EEG_β (12–30Hz)	EEG_γ (30–45Hz)	SleepStage
1	12.1	8.4	4.7	3.1	0.5	NREM1
2	18.2	6.2	2.3	1.2	0.3	NREM3
3	10.5	7.8	5.5	3.2	0.4	REM
4	17.9	5.1	2.0	1.1	0.2	NREM3
5	9.7	9.5	6.1	3.5	0.6	REM
6	5.4	11.2	8.3	4.9	1.2	Wake
7	14.1	7.0	4.4	2.8	0.3	NREM2
8	6.2	10.7	7.9	4.1	0.9	Wake
9	11.0	8.1	5.2	3.0	0.4	REM
10	13.3	6.5	3.9	2.2	0.2	NREM2

- DATASET 2

ID	Conn_1	Conn_2	Conn_3	Conn_4	Conn_5	Conn_6	Conn_7	Conn_8	Conn_9	Conn_10
1	0.81	0.64	0.12	-0.22	0.45	0.03	0.66	0.58	0.20	0.11
2	0.35	0.57	0.14	-0.18	0.52	0.07	0.59	0.50	0.18	0.09
3	0.72	0.62	0.10	-0.31	0.40	0.01	0.64	0.55	0.22	0.15
4	0.11	0.33	0.05	-0.12	0.60	0.08	0.55	0.47	0.12	0.02
5	0.90	0.79	0.20	-0.40	0.33	-0.02	0.69	0.61	0.28	0.19
6	0.77	0.68	0.13	-0.25	0.49	0.00	0.60	0.53	0.24	0.13
7	0.36	0.41	0.07	-0.15	0.57	0.05	0.58	0.48	0.16	0.10
8	0.80	0.70	0.17	-0.28	0.42	0.02	0.63	0.56	0.26	0.18
9	0.39	0.50	0.09	-0.20	0.55	0.06	0.57	0.49	0.20	0.11
10	0.85	0.73	0.18	-0.35	0.38	0.01	0.67	0.60	0.27	0.17

¿En cuál dataset se pueden aplicar métodos de Aprendizaje supervisado y en cuál método de Aprendizaje no supervisado?

Dataset 1: Supervisado, porque tiene la columna SleepStage, que es la etiqueta.

Dataset 2: No supervisado, porque solo tiene características y no tiene ninguna variable de salida.

3. ¿Cuál es la diferencia entre un problema de clasificación y uno de regresión?

Problema de clasificación: Se predicen categorías (spam/ no spam)

Regresión: Se predicen valores numéricos (Edad, peso)

4. ¿Cuál es la diferencia entre over-fitting y under-fitting?

Overfitting: El modelo aprende muy bien los datos de entrenamiento y no generaliza bien en datos nuevos. Tiene baja capacidad de predicción fuera del entrenamiento.

Underfitting: El modelo es muy simple y no captura la relación entre las variables, falla en los datos de entrenamiento y en los nuevos datos.

5. Observa la siguiente tabla con valores de accuracy en los dataset de entrenamiento y prueba

Ensayo	Accuracy Entrenamiento	Accuracy Prueba
1	99%	70%
2	60%	58%
3	65%	62%
4	95%	68%
5	92%	90%
6	100%	60%

Basado en los valores anteriores, ¿qué ensayo dirías que corresponde a casos con over-fitting, under-fitting o balanceado?

1. Overfitting
2. Underfitting
3. Underfitting
4. Overfitting
5. Balanceado
6. Overfitting

6. Contesta lo siguiente

A. Escribe las fórmulas de accuracy, precision, recall y f1-score

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

B. Con base a la siguiente tabla, calcula accuracy, precision, recall y f1-score

	Predicho Positivo	Predicho Negativo
Real Positivo	TP = 40	FN = 10
Real Negativo	FP = 5	TN = 45

$$\text{Accuracy} = (40 + 45) / (40 + 45 + 5 + 10) = 85 / 100 = 0.85$$

$$\text{Precision} = 40 / (40 + 5) = 40 / 45 = 0.88$$

$$\text{Recall} = 40 / (40 + 10) = 40 / 50 = 0.80$$

$$\text{F1-score} = 2 * (0.89 * 0.80) / (0.89 + 0.80) = 0.84$$

C. A continuación, tienes 3 descripciones de métricas en un problema de clasificación, indica a qué métrica corresponde

Descripción	Métrica
Proporción de predicciones correctas (tanto positivas como negativas) sobre el total de predicciones realizadas.	Accuracy
Porcentaje de verdaderos positivos identificados correctamente, respecto al total de positivos reales.	Recall
Porcentaje de aciertos entre los casos que el modelo predijo como positivos	Precisión

7. A continuación, se presentan varias características de algoritmos de Machine Learning supervisado. Indica a qué método corresponde cada una de las siguientes afirmaciones:

Nº	Característica	Método al que pertenece
1	Asume que los datos de cada clase siguen una distribución normal (gaussiana)	A) Análisis discriminantes
2	No tiene fase de entrenamiento como tal; toda la predicción ocurre en tiempo real	B) Vecino más cercano

3	Busca el hiperplano que maximiza el margen entre clases	C) Máquina de soporte vectorial
4	Realiza múltiples divisiones binarias para clasificar los datos	D) Árbol de decisión
5	Es sensible a la escala de las características, especialmente con distancia euclidiana	B) Vecino más cercano
6	Utiliza funciones núcleo (kernel) para trabajar con datos no lineales	C) Máquina de soporte vectorial
7	Calcula probabilidad bajo modelos estadísticos	A) Análisis discriminantes
8	Clasifica comparando con las más cercanas en el espacio de características	B) Vecino más cercano
9	Usa múltiples árboles para reducir varianza y mejorar precisión	E) Bosques aleatorios
10	Puede sobreajustar si el árbol es muy profundo	D) Árbol de decisión
11	Basado en modelos lineales con varianzas iguales por clase	A) Análisis discriminantes
12	Puede usar núcleos como RBF o polinomiales	C) Máquina de soporte vectorial
13	Mejora la generalización comparado con un solo árbol	E) Bosques aleatorios
14	Altamente interpretativo con pocas clases y atributos	D) Árbol de decisión
15	Muchos árboles sobre subconjuntos aleatorios	E) Bosques aleatorios

8. ¿Cuál es la diferencia entre Perceptron, Red Neuronal Artificial y Red Neuronal Convolutiva?

Perceptrón: Modelo más simple, tiene una sola capa y toma decisiones lineales. Es para problemas sencillos

Red neuronal artificial: Muchas capas y permite representar relaciones no lineales complejas

Red neuronal Convolutiva: Procesar datos como imágenes, utiliza capas convolucionales que extraen las características locales

9. En el contexto de aprendizaje de una red neuronal, ¿qué es un epoch?

Es una pasada completa por todo el conjunto de datos de entrenamiento, el modelo ajusta sus pesos comparando sus predicciones con respuestas reales y calculando el error. Usa los errores para mejorar

10. Lista las principales funciones de activación y describe la diferencia entre las funciones sigmoid y softmax

Rectified Linear Unit: Es una de las más utilizadas en redes profundas, es eficiente y permite que el modelo aprenda más rápido.

Sigmoid: Toma cualquier número real y lo convierte en un valor entre 0 y 1, se usa en la salida de redes para hacer clasificación binaria.

Tangente hiperbólica: Igual a sigmoid, pero sus valores van de -1 a 1, se usa cuando se quiere que las salidas estén centradas en 0

Softmax: Se usa en problemas de clasificación multiclase, transforma un vector de valores en una distribución de probabilidad donde todas las salidas suman 1

Diferencia entre sigmoid y softmax:

- **Sigmoid:** Actúa sobre una sola neurona, dando una probabilidad individual entre 0 y 1. Clasificación binaria.
- **Softmax:** se aplica a todas las salidas de la capa final, generando una probabilidad para cada clase. Todas esas probabilidades suman 1. Clasificación multiclase.