

Test de Repaso - Algoritmos de Machine Learning

1. ¿Qué función transforma salidas en probabilidad en regresión logística?

- A) ReLU
- B) Tangente hiperbólica
- C) Sigmoide
- D) Lineal

2. ¿Qué salida da un árbol de regresión?

- A) Clase más frecuente
- B) Valor promedio en la hoja
- C) Centroides del clúster
- D) Probabilidad de pertenencia

3. ¿Cuál es una forma de evitar el sobreajuste en árboles?

- A) Aumentar la profundidad
- B) Reducir el número de muestras
- C) Usar poda o limitar profundidad
- D) Eliminar variables

4. ¿Qué representa WCSS en K-means?

- A) Número de clústeres
- B) Precisión de la clasificación
- C) Error de clasificación
- D) Suma de distancias cuadradas a centroides

5. ¿Qué tipo de modelo es la regresión lineal?

- A) Paramétrico
- B) No supervisado
- C) No paramétrico
- D) Basado en vecinos

6. ¿Qué hace un pipeline en scikit-learn?

- A) Crea un gráfico
- B) Secuencia pasos de preprocesamiento y modelado
- C) Calcula WCSS
- D) Clasifica con árboles y redes

7. ¿Qué técnica mejora la generalización de árboles?

- A) Reducir K
- B) Random Forest
- C) Usar WCSS
- D) Aumentar profundidad

8. ¿Qué representa un F1-score alto?

- A) Recall bajo y precision alto
- B) Buena precisión pero mal recall

Test de Repaso - Algoritmos de Machine Learning

- C) Balance entre precisión y recall
 - D) Ninguno
9. ¿Qué parámetro en árboles limita el crecimiento del árbol?
- A) min_samples_leaf
 - B) n_clusters
 - C) max_depth
 - D) learning_rate
10. ¿Qué pasa si un modelo tiene accuracy alto pero recall bajo?
- A) Clasifica todo bien
 - B) Falla en detectar positivos reales
 - C) Detecta todos los positivos
 - D) Tiene pocos falsos positivos
11. ¿Qué tipo de modelo es KNN?
- A) Supervisado y paramétrico
 - B) No supervisado
 - C) No supervisado y paramétrico
 - D) Supervisado y no paramétrico
12. ¿Qué técnica usa validación cruzada?
- A) Para mejorar WCSS
 - B) Para prevenir overfitting
 - C) Para construir árboles
 - D) Para reducir el MAE
13. ¿Qué es un error residual?
- A) Valor predicho menos real
 - B) Distancia al centroide
 - C) Error absoluto
 - D) Valor real menos predicho
14. ¿Qué mide el recall en clasificación?
- A) Positivos reales detectados
 - B) Falsos positivos
 - C) Negativos predichos correctamente
 - D) Precisión media
15. ¿Qué hace la regresión logística?
- A) Predice valores continuos
 - B) Clasifica usando una función sigmoide
 - C) Agrupa observaciones
 - D) Calcula distancias a centroides

Test de Repaso - Algoritmos de Machine Learning

16. ¿Cuál de estas es una métrica para regresión?
- A) F1-score
 - B) Accuracy
 - C) MSE
 - D) Precision
17. ¿Cuál de estas métricas penaliza más los errores grandes?
- A) MAE
 - B) MSE
 - C) Accuracy
 - D) Recall
18. ¿Qué mide la métrica R^2 ?
- A) Error cuadrático
 - B) Varianza de residuos
 - C) Proporción de varianza explicada
 - D) Media de predicciones correctas
19. ¿Cuál es una limitación de MSE?
- A) No penaliza errores grandes
 - B) No es interpretable
 - C) Penaliza mucho los errores grandes
 - D) Requiere normalización
20. ¿Qué hace el método del codo?
- A) Calcula la profundidad óptima del árbol
 - B) Determina el mejor valor de K en KNN
 - C) Detecta el mejor número de clústeres
 - D) Ajusta coeficientes de regresión
21. ¿Cuál es una característica de los modelos paramétricos?
- A) Requieren todo el dataset para predecir
 - B) Aprenden un número fijo de parámetros
 - C) No hacen suposiciones sobre la forma
 - D) Usan clustering
22. ¿Qué hace RandomizedSearchCV?
- A) Genera nuevos datos
 - B) Optimiza hiperparámetros probando aleatoriamente
 - C) Reduce el overfitting
 - D) Agrupa observaciones similares
23. ¿Qué métrica usar si nos importan mucho los falsos negativos?
- A) Precision
 - B) Recall

Test de Repaso - Algoritmos de Machine Learning

- C) Accuracy
 - D) MAE
24. ¿Qué algoritmo clasifica observaciones calculando la distancia con los datos almacenados?
- A) Árboles de decisión
 - B) Regresión logística
 - C) KNN
 - D) K-means
25. ¿Qué significa sobreajuste (overfitting)?
- A) Modelo no aprende nada
 - B) Modelo memoriza datos de entrenamiento
 - C) Modelo generaliza bien
 - D) Modelo no tiene errores en prueba
26. ¿Qué mide la precisión en clasificación?
- A) Aciertos sobre todos los datos
 - B) Aciertos sobre los positivos reales
 - C) Aciertos entre los predichos como positivos
 - D) Aciertos entre los negativos reales
27. ¿Qué sucede si K en KNN es muy grande?
- A) El modelo se vuelve más preciso
 - B) Pierde flexibilidad y se suaviza demasiado
 - C) Se sobreajusta
 - D) Aumenta la variabilidad
28. ¿Qué mide la métrica MAE?
- A) Media de errores al cuadrado
 - B) Media del logaritmo de los errores
 - C) Media de los errores absolutos
 - D) Raíz cuadrada del error medio
29. ¿Qué pasa si no se escala bien un dataset en K-means?
- A) Nada relevante
 - B) Mejora el agrupamiento
 - C) Distorsiona los clústeres
 - D) Aumenta precisión
30. ¿Cuál es una ventaja del algoritmo KNN?
- A) Rápida predicción
 - B) Generaliza bien con pocos datos
 - C) Muy interpretable
 - D) Flexible y sin suposiciones fuertes