

Tema 1: Compresión de las bases del aprendizaje supervisado

1. ¿Cuál es el objetivo principal de una función de pérdida en aprendizaje supervisado?
 - a) Maximizar la cantidad de datos en el conjunto de entrenamiento
 - b) Minimizar la diferencia entre las predicciones y los valores reales
 - c) Aumentar la complejidad del modelo
 - d) Reducir el tamaño del conjunto de test

2. ¿Qué representa la curva de aprendizaje en un modelo de machine learning?
 - a) La evolución del error del modelo en diferentes épocas de entrenamiento
 - b) La velocidad con la que un modelo aprende nuevos datos
 - c) La cantidad de datos necesarios para alcanzar una precisión del 100%
 - d) La cantidad de memoria utilizada por el modelo durante el entrenamiento

3. ¿Cuál es la función del conjunto de validación en el entrenamiento de un modelo?
 - a) Evaluar el rendimiento del modelo antes de usarlo en datos reales
 - b) Ajustar los hiperparámetros del modelo sin sesgo
 - c) Generar más datos para mejorar la generalización
 - d) Reducir la sobrecarga computacional del entrenamiento

4. ¿Qué métrica es más adecuada para evaluar un modelo de clasificación binaria cuando las clases están desbalanceadas?
 - a) Precisión
 - b) Recall
 - c) Exactitud
 - d) Errores cuadráticos medios

5. ¿Cuál de los siguientes no es un tipo de conjunto de datos en aprendizaje supervisado?
- a) Conjunto de entrenamiento
 - b) Conjunto de validación
 - c) Conjunto de predicción
 - d) Conjunto de prueba

Tema 2: Distinción y aplicación de los métodos básicos de clasificación

6. ¿Cuál es la principal característica del algoritmo k-NN?
- a) Utiliza árboles de decisión para clasificar datos
 - b) Predice la etiqueta de un dato basándose en la mayoría de sus vecinos más cercanos
 - c) Construye múltiples modelos y promedia sus predicciones
 - d) Utiliza redes neuronales para mejorar la precisión
7. ¿Cuál es la ventaja principal de los árboles de decisión?
- a) Son resistentes al sobreajuste
 - b) Son fáciles de interpretar y visualizar
 - c) Siempre producen predicciones exactas
 - d) No requieren preprocesamiento de datos
8. ¿Qué técnica utiliza el algoritmo Random Forest para mejorar la precisión de las predicciones?
- a) Utiliza un único árbol de decisión con gran profundidad
 - b) Promedia las predicciones de múltiples árboles de decisión
 - c) Usa redes neuronales profundas para tomar decisiones
 - d) Emplea solo los datos más relevantes para entrenar el modelo
9. ¿Cuál es la principal ventaja de las Máquinas de Vectores de Soporte (SVM)?

- a) Funcionan mejor en conjuntos de datos extremadamente grandes
- b) Son robustas en problemas de alta dimensión
- c) No necesitan parámetros de ajuste
- d) Utilizan árboles de decisión para clasificar los datos

10. ¿Qué técnica utiliza SVM para clasificar datos no linealmente separables?

- a) Función de pérdida
- b) Método de k vecinos más cercanos
- c) Transformación del espacio con un kernel
- d) Normalización de datos

Tema 3: Distinción y aplicación de los métodos básicos de regresión

11. ¿Cuál es el principal supuesto de la regresión lineal?

- a) Existe una relación no lineal entre las variables independientes y la dependiente
- b) La variable dependiente debe ser categórica
- c) Existe una relación lineal entre las variables
- d) No requiere datos numéricos

12. ¿Qué diferencia a la regresión múltiple de la regresión lineal simple?

- a) La regresión múltiple utiliza múltiples variables independientes
- b) La regresión múltiple solo se usa en redes neuronales
- c) No existen diferencias, son el mismo método
- d) La regresión múltiple no necesita entrenamiento

13. ¿En qué tipo de problemas se usa la regresión logística?

- a) Predicción de valores continuos
- b) Clasificación binaria
- c) Series temporales
- d) Modelado de redes neuronales

Tema 4: Herramientas de desarrollo de redes neuronales

14. ¿Qué es TensorFlow?

- a) Un framework de desarrollo para modelos de machine learning
- b) Un lenguaje de programación específico para redes neuronales
- c) Una técnica de reducción de dimensionalidad
- d) Un algoritmo de regresión logística

15. ¿Cuál es una aplicación común de redes neuronales en visión por computadora?

- a) Clasificación de imágenes
- b) Predicción de series temporales
- c) Cálculo de estadísticas descriptivas
- d) Algoritmos de búsqueda

Tema 5: Implementación y puesta en marcha

16. ¿Para qué se usa Docker en aprendizaje automático?

- a) Para gestionar contenedores de software que contienen modelos entrenados
- b) Para entrenar modelos de deep learning
- c) Para limpiar datos antes del entrenamiento
- d) Para reducir el tiempo de entrenamiento de redes neuronales

Tema 6: Ética y limitaciones

17. ¿Qué problema puede surgir al entrenar un modelo con datos sesgados?

- a) El modelo puede discriminar ciertos grupos de manera injusta
- b) El modelo aprende más rápido

- c) Se reduce la complejidad del modelo
- d) Se mejora la eficiencia computacional

Tema 1: Compresión de las bases del aprendizaje supervisado

18. ¿Cuál es una señal clara de sobreajuste (overfitting) en un modelo de machine learning?

- a) Alto rendimiento en los datos de entrenamiento, pero bajo rendimiento en los de prueba
- b) Rendimiento uniforme en entrenamiento y prueba, pero con alta precisión
- c) Bajo rendimiento tanto en entrenamiento como en prueba
- d) Reducción de la pérdida en todos los conjuntos de datos

19. ¿Cuál de las siguientes métricas es más adecuada para evaluar un modelo de regresión?

- a) F1 Score
- b) Matriz de confusión
- c) Error cuadrático medio (MSE)
- d) Precisión

20. ¿Por qué es importante dividir los datos en conjuntos de entrenamiento, validación y prueba?

- a) Para evitar sesgos en el entrenamiento y evaluar la generalización del modelo
- b) Para aumentar la precisión del modelo
- c) Para mejorar la eficiencia computacional
- d) Para crear modelos más complejos

Tema 2: Distinción y aplicación de los métodos básicos de clasificación

21. ¿Cuál es una desventaja del algoritmo k-NN?
- a) Es difícil de interpretar
 - b) Requiere una fase de entrenamiento muy larga
 - c) Su rendimiento disminuye con conjuntos de datos muy grandes
 - d) No puede manejar datos categóricos
22. En un árbol de decisión, ¿qué criterio se usa comúnmente para dividir los nodos?
- a) Regresión lineal
 - b) Información mutua
 - c) Entropía o índice de Gini
 - d) Curva ROC
23. ¿Cuál es una ventaja clave de los Random Forests frente a un único árbol de decisión?
- a) Menor necesidad de preprocesamiento de datos
 - b) Mayor interpretabilidad del modelo
 - c) Reducción del sobreajuste mediante el promedio de múltiples árboles
 - d) Mayor velocidad de entrenamiento
24. ¿Qué parámetro es crucial para el rendimiento de una SVM?
- a) Número de vecinos
 - b) Función de kernel
 - c) Profundidad del árbol
 - d) Número de capas ocultas

Tema 3: Distinción y aplicación de los métodos básicos de regresión (Continuación)

25. ¿Cuál es la diferencia principal entre regresión lineal y regresión logística?
- a) La regresión logística se usa para valores continuos y la lineal para clasificación
 - b) La regresión lineal predice valores continuos y la logística probabilidades de clase
 - c) No hay diferencias entre ellas
 - d) La regresión logística solo usa una variable independiente
26. ¿En qué situación usarías una regresión múltiple?
- a) Cuando hay más de una variable independiente
 - b) Para clasificar datos binarios
 - c) Cuando los datos están desbalanceados
 - d) Para reducir la dimensionalidad de los datos
27. ¿Qué técnica ayuda a evitar el sobreajuste en modelos de regresión?
- a) Regularización (Lasso o Ridge)
 - b) Aumento del número de variables independientes
 - c) Uso exclusivo del conjunto de entrenamiento
 - d) Eliminación del conjunto de validación

Tema 4: Herramientas de desarrollo de redes neuronales

28. ¿Qué tipo de red neuronal es más adecuada para visión por computadora?
- a) Redes neuronales recurrentes (RNN)
 - b) Redes neuronales convolucionales (CNN)
 - c) Redes neuronales densas
 - d) Máquinas de vectores de soporte
29. ¿Qué ventaja ofrece TensorFlow para el desarrollo de modelos de machine learning?

- a) Simplicidad en el entrenamiento de modelos pequeños
- b) Integración con lenguajes de bajo nivel como C++
- c) Escalabilidad y soporte para procesamiento en GPU
- d) Menor consumo de memoria durante el entrenamiento

30. ¿Cuál es un caso de uso típico de redes neuronales en procesamiento de lenguaje natural (NLP)?

- a) Clasificación de imágenes
- b) Predicción de valores continuos
- c) Análisis de sentimientos en textos
- d) Detección de anomalías en sensores

Tema 5: Implementación y puesta en marcha (Continuación)

31. ¿Por qué se usa Docker en el despliegue de modelos de machine learning?

- a) Para asegurar la reproducibilidad del entorno de ejecución
- b) Para entrenar redes neuronales más rápido
- c) Para realizar visualización de datos
- d) Para limpiar y transformar datasets

32. ¿Cuál es una ventaja de desplegar modelos en servicios en la nube?

- a) Reducción de la precisión del modelo
- b) Mayor tiempo de respuesta
- c) Escalabilidad automática de recursos
- d) Eliminación de la necesidad de entrenamiento

Tema 6: Ética y limitaciones (Continuación)

33. ¿Por qué es importante considerar el sesgo en los datos de entrenamiento?

- a) Para mejorar el rendimiento del modelo en el conjunto de entrenamiento

- b) Para evitar decisiones injustas o discriminatorias
- c) Para aumentar la complejidad del modelo
- d) Para reducir el tiempo de entrenamiento

34. ¿Cuál de los siguientes es un desafío ético en el uso de IA?

- a) Reducción del tiempo de entrenamiento
- b) Interpretabilidad de los modelos complejos
- c) Incremento de la precisión
- d) Optimización del hardware

35. ¿Qué aspecto ético está relacionado con la gestión de la incertidumbre en las decisiones automatizadas?

- a) Transparencia en la toma de decisiones
- b) Reducción del tiempo de cómputo
- c) Aumento del rendimiento en producción
- d) Minimización de la pérdida

Tema 5: Implementación y puesta en marcha de un sistema de aprendizaje automático

36. ¿Cuál es una práctica recomendada al exponer un modelo de machine learning mediante una API?

- a) Usar formatos propietarios para el intercambio de datos
- b) Implementar endpoints claros y bien documentados
- c) Entrenar el modelo dentro de la API en cada solicitud
- d) Desplegar el modelo sin realizar pruebas previas

37. ¿Cuál es un beneficio clave de usar contenedores como Docker en machine learning?

- a) Mejora la precisión del modelo
- b) Aumenta la velocidad de entrenamiento
- c) Facilita la portabilidad y consistencia entre entornos
- d) Reduce la cantidad de datos necesarios para el entrenamiento

38. ¿Qué componente es esencial en la gestión del ciclo de vida de un modelo de machine learning?

- a) Visualización de datos únicamente
- b) Monitoreo del rendimiento del modelo en producción
- c) Despliegue sin validación previa
- d) Uso exclusivo de modelos estáticos

Tema 6: Identificación de las limitaciones de los sistemas de aprendizaje y de las consecuencias éticas de su mal uso

39. ¿Cuál es un ejemplo de sesgo en un modelo de machine learning?

- a) Un modelo que aprende rápidamente en todas las muestras
- b) Un modelo que favorece a ciertos grupos debido a un desequilibrio en los datos
- c) Una precisión del 100% en los datos de entrenamiento
- d) Un modelo que siempre predice la misma clase

40. ¿Cuál es una posible consecuencia ética del uso irresponsable de sistemas de aprendizaje automático?

- a) Reducción del tiempo de entrenamiento
- b) Aumento de la eficiencia en el procesamiento de datos
- c) Discriminación o falta de equidad en las decisiones automatizadas
- d) Incremento en la capacidad de almacenamiento de datos