

Árboles, Regresión y SVM: Test

Opción Múltiple:

1. ¿Cuál de las siguientes técnicas de aprendizaje automático permite identificar conceptos a partir de las características de un conjunto de ejemplos organizados jerárquicamente?
 - a) Regresión Lineal
 - b) Árboles de Decisión
 - c) Regresión Logística
 - d) Máquinas de Vectores de Soporte
2. ¿Qué representan las hojas en un árbol de decisión?
 - a) Preguntas o tests sobre los ejemplos
 - b) Nodos intermedios
 - c) Los conceptos extraídos o una clase específica
 - d) Ramas del árbol
3. ¿Cuál de los siguientes criterios de selección de separadores busca la menor impureza en un nodo?
 - a) Entropía
 - b) Ganancia de Información
 - c) Índice de Gini
 - d) Tasa de ramificación
4. ¿Qué técnica de conjunto de modelos crea múltiples árboles a partir de muestras aleatorias del conjunto de datos de entrenamiento con reemplazo?
 - a) Boosting
 - b) Bagging
 - c) Random Forests
 - d) Boosted Trees
5. ¿Qué algoritmo de conjunto de modelos construye secuencialmente árboles de decisión, donde cada árbol intenta corregir los errores del anterior?
 - a) Bagging
 - b) Random Forests
 - c) Boosting
 - d) Promediado simple

6. ¿Cuál de las siguientes técnicas de optimización de hiperparámetros prueba exhaustivamente todas las combinaciones de un conjunto de valores especificados?
- a) Búsqueda Aleatoria
 - b) Búsqueda Exhaustiva (GridSearch)
 - c) Validación Cruzada
 - d) Descenso de Gradiente
7. En regresión lineal simple, ¿qué representan β_0 y β_1 ?
- a) Los errores del modelo
 - b) Las observaciones de entrada y salida
 - c) La intersección y la pendiente, respectivamente
 - d) El número de observaciones
8. ¿Qué enfoque se utiliza comúnmente para estimar los coeficientes en la regresión lineal minimizando la suma de los cuadrados de las diferencias entre los valores observados y predichos?
- a) Máxima Verosimilitud
 - b) Descenso por Gradiente
 - c) Descomposiciones Matriciales
 - d) Mínimos Cuadrados Ordinarios (MCO)
9. ¿Qué función se utiliza en la regresión logística para modelar la probabilidad de pertenencia a una clase, asegurando que la salida esté entre 0 y 1?
- a) Función lineal
 - b) Función sigmoide o logística
 - c) Función de pérdida de entropía cruzada
 - d) Función kernel
10. En las Máquinas de Vectores de Soporte (SVM), ¿qué son los vectores de soporte?
- a) Todos los datos de entrenamiento
 - b) Los datos que están lejos del hiperplano
 - c) Las observaciones que se encuentran en el margen o lo violan
 - d) Los hiperplanos de separación

11. ¿Cuál es el objetivo principal de la regularización en regresión lineal?

- a) Aumentar la complejidad del modelo
- b) Evitar el sobreajuste
- c) Mejorar la interpretabilidad de los datos
- d) Acelerar el entrenamiento del modelo

12. ¿Qué tipo de regularización en regresión lineal puede llevar algunos coeficientes a ser exactamente cero, realizando selección de características?

- a) Regularización L2 (Ridge)
- b) Regularización L1 (Lasso)
- c) Elastic Net
- d) Sin regularización

13. ¿Qué representa el sesgo (bias) en el error de un modelo de Machine Learning?

- a) La sensibilidad del modelo a pequeños cambios en los datos de entrenamiento
- b) La diferencia entre el valor proporcionado por el modelo y el valor real (error)
- c) La cantidad de variación de la predicción del modelo según los datos de entrenamiento
- d) El error en los datos de entrenamiento

14. ¿Qué representa la varianza en el error de un modelo de Machine Learning?

- a) La capacidad del modelo para aprender los patrones de los datos de entrenamiento
- b) La diferencia entre el valor proporcionado por el modelo y el valor real (error)
- c) La cantidad de variación de la predicción del modelo según los datos que se usen para el entrenamiento
- d) El error en los datos de prueba

15. ¿Cuál es el objetivo del algoritmo de descenso del gradiente?

- a) Encontrar la derivada de una función
- b) Encontrar el mínimo de una función de coste
- c) Calcular el error cuadrático medio
- d) Normalizar los datos de entrada

Verdadero/Falso:

16. Un árbol de decisión siempre realiza una clasificación perfecta de los datos de entrenamiento.
17. La poda de un árbol de decisión busca reducir el sobreajuste.
18. En Bagging, cada árbol individual se entrena con el mismo subconjunto de datos.
19. Random Forests añade una capa de aleatoriedad al Bagging al seleccionar un subconjunto de características en cada división.
20. Boosting busca principalmente reducir la varianza del modelo.
21. Los hiperparámetros de un modelo se aprenden directamente de los datos de entrenamiento.
22. La regresión lineal asume una relación lineal entre las variables predictoras y la variable objetivo.
23. El método de mínimos cuadrados promedio (LMS) se utiliza cuando se dispone de un gran número de datos medidos.
24. La regresión logística se utiliza para predecir valores continuos.
25. En regresión logística, un coeficiente negativo indica una correlación positiva con la clase predicha.
26. Las Máquinas de Vectores de Soporte (SVM) funcionan bien con datos de baja dimensión.
27. Un kernel lineal en SVM puede separar datos que no son linealmente separables en el espacio original.
28. La regularización siempre mejora el rendimiento de un modelo lineal.
29. Escalar las variables predictoras es siempre necesario al usar regresión lineal sin regularización.
30. Un modelo con alto sesgo tiende a sobreajustar los datos de entrenamiento.
31. Un modelo con alta varianza tiende a generalizar bien a datos nuevos.
32. El objetivo ideal es un modelo con bajo sesgo y baja varianza.
33. El ratio de aprendizaje en el descenso del gradiente siempre debe ser grande para asegurar una convergencia rápida.
34. Una función convexa garantiza que cualquier mínimo local encontrado será también el mínimo global.
35. El gradiente de una función multidimensional apunta en la dirección de mayor descenso.

Respuesta Corta:

36. ¿Cuál es el primer paso en el algoritmo general para construir un árbol de decisión?
37. Menciona dos criterios de parada comunes al construir un árbol de decisión para evitar el sobreaprendizaje.
38. ¿Qué significa "pruning" en el contexto de los árboles de decisión?
39. ¿Cómo se combinan las predicciones de los árboles individuales en Bagging para obtener la predicción final?
40. ¿Qué parámetro clave distingue a Random Forests de Bagging?
41. En Boosted Trees, ¿qué intenta predecir cada árbol en lugar del campo objetivo?
42. Define brevemente qué es un hiperparámetro.
43. Menciona dos métodos genéricos para la búsqueda de la mejor combinación de hiperparámetros para un modelo.
44. ¿Qué mide la Suma del Error Cuadrático (RSS) en regresión lineal?
45. ¿Cuál es el supuesto principal detrás del enfoque de Máxima Verosimilitud para estimar los parámetros en regresión logística?
46. ¿Qué problema principal buscan resolver los kernels en las Máquinas de Vectores de Soporte (SVM)?
47. Menciona un tipo común de función kernel utilizada en SVM.
48. ¿Cuál es la principal desventaja de los árboles de decisión individuales?
49. ¿Qué se busca al realizar la poda posterior (post-pruning) de un árbol de decisión?
50. ¿Qué indica un valor muy pequeño del ratio de aprendizaje en el descenso del gradiente?

Soluciones

Respuesta Múltiple

1. b
2. c
3. c
4. b
5. c
6. b
7. c
8. d
9. b
10. c
11. b
12. b
13. b
14. c
15. b

Verdadero/Falso

16. Falso
17. Verdadero
18. Falso
19. Verdadero
20. Falso
21. Falso
22. Verdadero
23. Falso
24. Falso
25. Falso
26. Falso
27. Falso
28. Falso
29. Falso
30. Falso
31. Falso
32. Verdadero
33. Falso
34. Verdadero
35. Falso

Respuesta Corta

36. **Calcular la calidad del nodo (conjunto de ejemplos)**
37. **Máximo número de ramas por nodo, número mínimo de observaciones por nodo final, número mínimo de observaciones para dividir un nodo, variables discriminantes**
(menciona dos cualesquiera)
38. **Reducir el árbol, haciéndolo más sencillo eliminando nodos redundantes**
39. **Se promedian (para regresión) o se realiza votación por mayoría (para clasificación)**
40. **La selección aleatoria de un subconjunto de las características en cada división**
41. **Los errores cometidos por el modelo anterior**
42. **Un parámetro externo al modelo que es fijado arbitrariamente por el programador antes del entrenamiento**
43. **Búsqueda exhaustiva (GridSearch) y Búsqueda aleatoria (Random Search)**
44. **La suma de los cuadrados de las diferencias entre los valores observados y los predichos**
45. **Buscar los valores de los parámetros que hagan que los resultados predichos sean lo más consistentes posible con los datos observados (o similar)**
46. **Permitir operar en espacios de alta dimensionalidad de manera eficiente sin realizar cálculos explícitos en ese espacio**
47. **Kernel polinomial, Kernel de base radial (RBF) (menciona uno)**
48. **Son propensos al sobreajuste**
49. **Simplificar un árbol completamente construido para mejorar la generalización y reducir el sobreajuste**
50. **Pasos muy pequeños en cada iteración, lo que puede llevar a una convergencia lenta o quedar atrapado en mínimos locales**