

Área personal

Mis cursos

Sistemas de Aprendizaje Automático

3. Algoritmos aplicados al aprendizaje supervisado y optimización del modelo

Cuestionario "Chapter 2. End-to-End Machine Learning Project"

Comenzado el	miércoles, 5 de febrero de 2025, 18:46
Estado	Finalizado
Finalizado en	miércoles, 5 de febrero de 2025, 19:11
Tiempo empleado	24 minutos 41 segundos
Puntos	24,00/30,00
Calificación	8,00 de 10,00 (80%)

Pregunta 1

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué significa el concepto de "monitorización" en un modelo en producción?

Seleccione una:

- ☒ a. Revisar su desempeño con datos nuevos regularmente ✓
- ☐ b. Volver a entrenarlo cada día
- ☐ c. Cambiar manualmente los pesos del modelo
- ☐ d. Descartar modelos anteriores
- ☐ e. Todas son falsas

La respuesta correcta es: Revisar su desempeño con datos nuevos regularmente

Pregunta 2

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué es lo primero que hay que hacer en un proyecto de Machine Learning?

Seleccione una:

- ☒ a. Averiguar cuál es el objetivo principal del proyecto ✓
- ☐ b. Obtener los datos
- ☐ c. Seleccionar un modelo
- ☐ d. Entrenar el modelo
- ☐ e. Todas son falsas

La respuesta correcta es: Averiguar cuál es el objetivo principal del proyecto

Pregunta 3

Sin contestar

Se puntúa como 0 sobre 1,00

¿Qué estrategia de ensamble combina modelos entrenados con diferentes subconjuntos del conjunto de entrenamiento?

Seleccione una:

- ☐ a. Bagging
- ☐ b. Stacking
- ☐ c. Boosting
- ☐ d. Dropout
- ☐ e. Todas son falsas

La respuesta correcta es: Bagging

Pregunta 4

Correcta

Se puntúa 1,00 sobre 1,00

¿Por qué es importante evaluar continuamente un modelo en producción?

Seleccione una:

- ☒ a. Para detectar posibles degradaciones en su desempeño ✓
- ☐ b. Porque los modelos nunca cambian su precisión
- ☐ c. Para evitar entrenar modelos nuevos
- ☐ d. Porque los datos de entrada siempre son iguales
- ☐ e. Todas son falsas

La respuesta correcta es: Para detectar posibles degradaciones en su desempeño

Pregunta 5

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál de las siguientes fuentes NO es mencionada en el capítulo para obtener datos abiertos?

Seleccione una:

- ☒ a. Google Play Store ✓
- ☐ b. OpenML
- ☐ c. Kaggle
- ☐ d. Amazon AWS Datasets
- ☐ e. Todas son falsas

La respuesta correcta es: Google Play Store

Pregunta 6

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué herramienta de Scikit-Learn facilita la transformación de datos en una secuencia automatizada?

Seleccione una:

- ☒ a. Pipeline ✓
- ☐ b. GridSearchCV
- ☐ c. SelectKBest
- ☐ d. Ridge Regression
- ☐ e. Todas son falsas

La respuesta correcta es: Pipeline

Pregunta 7

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué se debe hacer cuando se detecta un cambio en la distribución de los datos en un modelo en producción?

Seleccione una:

- ☒ a. Reentrenar el modelo con datos actualizados ✓
- ☐ b. Ignorar el cambio y seguir usando el modelo
- ☐ c. Eliminar todos los datos históricos
- ☐ d. Dejar de monitorear el modelo
- ☐ e. Todas son falsas

La respuesta correcta es: Reentrenar el modelo con datos actualizados

Pregunta 8

Correcta

Se puntúa 1,00 sobre 1,00

¿Por qué es importante almacenar versiones anteriores de los modelos?

Seleccione una:

- ☒ a. Para poder restaurarlos si ocurre un problema ✓
- ☐ b. Para reducir el tamaño del dataset
- ☐ c. Para hacer el entrenamiento más lento
- ☐ d. Para evitar probar nuevos modelos
- ☐ e. Todas son falsas

La respuesta correcta es: Para poder restaurarlos si ocurre un problema

Pregunta 9

Correcta

Se puntúa 1,00 sobre 1,00

¿Por qué es importante evaluar un modelo con datos de prueba antes de desplegarlo?

Seleccione una:

- ☒ a. Para asegurarse de que el modelo generaliza bien ✓
- ☐ b. Para aumentar la precisión en el conjunto de entrenamiento
- ☐ c. Para evitar que los datos sean demasiado grandes
- ☐ d. Para reducir la velocidad de predicción
- ☐ e. Todas son falsas

La respuesta correcta es: Para asegurarse de que el modelo generaliza bien

Pregunta 10

Correcta

Se puntúa 1,00 sobre 1,00

¿Por qué es importante almacenar versiones anteriores de los modelos?

Seleccione una:

- ☒ a. Para poder restaurarlos si ocurre un problema ✓
- ☐ b. Para reducir el tamaño del dataset
- ☐ c. Para hacer el entrenamiento más lento
- ☐ d. Para evitar probar nuevos modelos
- ☐ e. Todas son falsas

La respuesta correcta es: Para poder restaurarlos si ocurre un problema

Pregunta 11

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál de las siguientes técnicas se usa para manejar valores faltantes en un conjunto de datos?

Seleccione una:

- ☒ a. Todas las anteriores ✓
- ☐ b. Eliminar las filas
- ☐ c. Llenar los valores con la mediana
- ☐ d. Usar modelos de imputación
- ☐ e. Todas son falsas

La respuesta correcta es: Todas las anteriores

Pregunta 12

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué método se usa para buscar automáticamente los mejores hiperparámetros de un modelo?

Seleccione una:

- ☒ a. Grid Search ✓
- ☐ b. One-hot encoding
- ☐ c. StandardScaler
- ☐ d. Regresión Ridge
- ☐ e. Todas son falsas

La respuesta correcta es: Grid Search

Pregunta 13

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué métrica se menciona como importante para evaluar modelos de regresión?

Seleccione una:

- ☒ a. RMSE (Root Mean Squared Error) ✓
- ☐ b. Precisión
- ☐ c. Sensibilidad
- ☐ d. AUC-ROC
- ☐ e. Todas son falsas

La respuesta correcta es: RMSE (Root Mean Squared Error)

Pregunta 14

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué tipo de modelo puede ayudar a manejar cambios en la distribución de datos con el tiempo?

Seleccione una:

- ☒ a. Modelos en línea (online learning) ✓
- ☐ b. Modelos estáticos
- ☐ c. Modelos sin ajuste de parámetros
- ☐ d. Modelos que no necesitan datos nuevos
- ☐ e. Todas son falsas

La respuesta correcta es: Modelos en línea (online learning)

Pregunta 15

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué modelo de Machine Learning se prueba primero en el capítulo?

Seleccione una:

- ☒ a. Regresión lineal ✓
- ☐ b. Random Forest
- ☐ c. Support Vector Machines
- ☐ d. Redes neuronales
- ☐ e. Todas son falsas

La respuesta correcta es: Regresión lineal

Pregunta 16

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál es una buena práctica cuando se presentan los resultados de un modelo?

Seleccione una:

- ☒ a. Comparar múltiples métricas de desempeño ✓
- ☐ b. Usar solo métricas de entrenamiento
- ☐ c. Ocultar información de evaluación
- ☐ d. No incluir análisis de errores
- ☐ e. Todas son falsas

La respuesta correcta es: Comparar múltiples métricas de desempeño

Pregunta 17

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué se recomienda hacer antes de desplegar un modelo en producción?

Seleccione una:

- ☒ a. Realizar pruebas de rendimiento ✓
- ☐ b. Ajustar hiperparámetros en el conjunto de prueba
- ☐ c. Desplegarlo sin probarlo
- ☐ d. Modificar manualmente los pesos del modelo
- ☐ e. Todas son falsas

La respuesta correcta es: Realizar pruebas de rendimiento

Pregunta 18

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál es el principal problema con un modelo que tiene un error bajo en el conjunto de entrenamiento pero un error alto en el conjunto de prueba?

Seleccione una:

- ☒ a. Sobreajuste ✓
- ☐ b. Subajuste
- ☐ c. Datos no balanceados
- ☐ d. Baja varianza
- ☐ e. Todas son falsas

La respuesta correcta es: Sobreajuste

Pregunta 19

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál de los siguientes es un problema que puede surgir después del despliegue de un modelo?

Seleccione una:

- ☒ a. Desviación en la distribución de datos ✓
- ☐ b. La precisión del modelo siempre mejora
- ☐ c. Los datos de entrada nunca cambian
- ☐ d. El modelo no necesita actualización
- ☐ e. Todas son falsas

La respuesta correcta es: Desviación en la distribución de datos

Pregunta 20

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué tipo de problema puede ocurrir si un modelo aprende demasiado de los datos de entrenamiento?

Seleccione una:

- ☒ a. Sobreajuste ✓
- ☐ b. Subajuste
- ☐ c. Reducción de dimensionalidad
- ☐ d. Regularización
- ☐ e. Todas son falsas

La respuesta correcta es: Sobreajuste

Pregunta 21

Incorrecta

Se puntúa -0,33 sobre 1,00

¿Por qué es importante separar los datos en conjunto de entrenamiento y conjunto de prueba?

Seleccione una:

- ☐ a. Para evitar sobreajuste
- ☐ b. Para mejorar la velocidad del modelo
- ☐ c. Para reducir la cantidad de datos necesarios
- ☐ d. Para mejorar la complejidad del modelo
- ☒ e. Todas son falsas ✗

La respuesta correcta es: Para evitar sobreajuste

Pregunta 22

Sin contestar

Se puntúa como 0 sobre 1,00

¿Qué técnica es útil para reducir el sobreajuste en modelos de Machine Learning?

Seleccione una:

- ☐ a. Aplicar regularización
- ☐ b. Usar conjuntos de datos más pequeños
- ☐ c. No usar validación cruzada
- ☐ d. Aumentar la cantidad de hiperparámetros
- ☐ e. Todas son falsas

La respuesta correcta es: Aplicar regularización

Pregunta 23

Incorrecta

Se puntúa -0,33 sobre 1,00

¿Qué técnica se recomienda para evaluar correctamente el rendimiento de un modelo en el entrenamiento?

Seleccione una:

- ☐ a. Validación cruzada
- ☐ b. Usar solo el conjunto de entrenamiento
- ☒ c. Ajustar hiperparámetros manualmente ✖
- ☐ d. Agregar más datos sin evaluar
- ☐ e. Todas son falsas

La respuesta correcta es: Validación cruzada

Pregunta 24

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué técnica es útil para escalar atributos numéricos en Machine Learning?

Seleccione una:

- ☒ a. Min-max scaling ✔
- ☐ b. One-hot encoding
- ☐ c. Reemplazo de valores faltantes
- ☐ d. Generación de histogramas
- ☐ e. Todas son falsas

La respuesta correcta es: Min-max scaling

Pregunta 25

Incorrecta

Se puntúa -0,33 sobre 1,00

¿Cuál es una ventaja de usar RandomizedSearchCV sobre GridSearchCV?

Seleccione una:

- ☐ a. Permite explorar más combinaciones en menos tiempo
- ☐ b. Siempre encuentra los mejores hiperparámetros
- ☐ c. No requiere datos de prueba
- ☐ d. No usa aleatoriedad
- ☒ e. Todas son falsas ✖

La respuesta correcta es: Permite explorar más combinaciones en menos tiempo

Pregunta 26

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué es la fuga de datos (data leakage) al usar el conjunto de prueba y cuáles son sus consecuencias?

Seleccione una:

- ☒ a. Ocurre cuando la información del conjunto de prueba influye en el entrenamiento del modelo, causando una evaluación artificialmente alta y un mal desempeño en datos reales. ✔
- ☐ b. Es cuando se pierden datos del conjunto de entrenamiento debido a errores en el preprocesamiento.
- ☐ c. Se produce cuando el modelo es entrenado con datos insuficientes y tiene un alto sesgo.
- ☐ d. Es un tipo de regularización aplicada para evitar el sobreajuste.
- ☐ e. Todas son falsas

La respuesta correcta es: Ocurre cuando la información del conjunto de prueba influye en el entrenamiento del modelo, causando una evaluación artificialmente alta y un mal desempeño en datos reales.

Pregunta 27

Correcta

Se puntúa 1,00 sobre 1,00

¿Qué es un Pipeline en Machine Learning y cuál es su principal utilidad?

Seleccione una:

- ☒ a. Es una secuencia de transformaciones de datos que facilita la automatización del preprocesamiento y modelado. ✔
- ☐ b. Es un algoritmo de optimización utilizado en redes neuronales profundas.
- ☐ c. Es un método para dividir datos en conjuntos de entrenamiento y prueba.
- ☐ d. Es una técnica utilizada únicamente en regresión logística.
- ☐ e. Todas son falsas

La respuesta correcta es: Es una secuencia de transformaciones de datos que facilita la automatización del preprocesamiento y modelado.

Pregunta 28

Correcta

Se puntúa 1,00 sobre 1,00

¿En qué caso es más recomendable usar Ordinal Encoder en vez de One-Hot Encoder?

Seleccione una:

- ☒ a. Cuando las categorías tienen un orden natural y la relación ordinal es relevante para el modelo. ✔
- ☐ b. Cuando las categorías son nominales y no tienen relación de orden.
- ☐ c. Cuando se desea reducir la dimensionalidad en cualquier tipo de variable categórica.
- ☐ d. Cuando el conjunto de datos solo contiene valores numéricos.
- ☐ e. Todas son falsas

La respuesta correcta es: Cuando las categorías tienen un orden natural y la relación ordinal es relevante para el modelo.

Pregunta **29**

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál es la principal diferencia entre una SciPy Sparse Matrix y un NumPy Array?

Seleccione una:

- ☒ a. Una SciPy Sparse Matrix almacena solo los elementos no nulos, reduciendo el consumo de memoria, mientras que un NumPy Array almacena todos los elementos, incluyendo ceros.
- ☐ b. Un NumPy Array es más eficiente en términos de memoria que una SciPy Sparse Matrix.
- ☐ c. Las SciPy Sparse Matrices solo pueden utilizarse en modelos de clasificación.
- ☐ d. Las SciPy Sparse Matrices son exclusivamente para datos de texto.
- ☐ e. Todas son falsas



La respuesta correcta es: Una SciPy Sparse Matrix almacena solo los elementos no nulos, reduciendo el consumo de memoria, mientras que un NumPy Array almacena todos los elementos, incluyendo ceros.

Pregunta **30**

Correcta

Se puntúa 1,00 sobre 1,00

¿Cuál es la forma correcta de aplicar un escalador (Scaler) en Machine Learning?

Seleccione una:

- ☒ a. Se debe ajustar (fit) el escalador solo con los datos de entrenamiento y luego aplicarlo (transform) al conjunto de validación, prueba y datos nuevos.
- ☐ b. Se puede usar fit() y fit\_transform() en cualquier conjunto de datos sin importar si es de entrenamiento o prueba.
- ☐ c. Los valores transformados siempre se mantienen dentro del mismo rango, sin importar si hay valores atípicos en datos nuevos.
- ☐ d. No es necesario usar escaladores en modelos de Machine Learning.
- ☐ e. Todas son falsas



La respuesta correcta es: Se debe ajustar (fit) el escalador solo con los datos de entrenamiento y luego aplicarlo (transform) al conjunto de validación, prueba y datos nuevos.

Actividad previa

◀ Foro de dudas

Ir a...

Siguiente actividad

Modelos basados en árboles de decisión ▶