

Test de Repaso - Preprocesamiento y Validación de Modelos

1. ¿Qué función asegura que los mismos datos se usen en splits aleatorios?
 - A) shuffle=False
 - B) seed=42
 - C) random_state=42
 - D) repeat=10
2. ¿Qué estrategia de imputación es más robusta frente a valores extremos?
 - A) Media
 - B) Moda
 - C) Mediana
 - D) Zero
3. ¿Qué hace el parámetro stratify en train_test_split?
 - A) Asegura aleatoriedad total
 - B) Equilibra la clase objetivo
 - C) Evita duplicados
 - D) Desactiva el shuffle
4. ¿Qué parámetro controla el número de particiones en cross_val_score?
 - A) split
 - B) fold
 - C) cv
 - D) repeat
5. ¿Qué método busca combinaciones aleatorias de hiperparámetros?
 - A) RandomTuner
 - B) RandomSearchCV
 - C) RandomGrid
 - D) FastCV
6. ¿Qué función permite construir un pipeline sin nombrar los pasos?
 - A) build_pipeline
 - B) make_pipeline
 - C) fast_pipeline
 - D) create_pipeline
7. ¿Qué función se usa para obtener el mejor modelo en GridSearchCV?
 - A) get_best_model
 - B) best_estimator_
 - C) get_estimator
 - D) select_model
8. ¿Qué peligro representa el data leakage?
 - A) Mejor rendimiento en test
 - B) Peor interpretación de coeficientes

Test de Repaso - Preprocesamiento y Validación de Modelos

- C) Predicciones triviales
 - D) Entrenamiento con información del test
9. ¿Qué método define un objeto con pasos secuenciales de transformación y modelado?
- A) Pipeline
 - B) Workflow
 - C) SequenceModel
 - D) EstimatorChain
10. ¿Qué codificador se usa si el orden de las categorías es importante?
- A) LabelEncoder
 - B) OneHotEncoder
 - C) OrdinalEncoder
 - D) CategoryEncoder
11. ¿Qué tipo de transformador es SimpleImputer?
- A) Estimador
 - B) Predicador
 - C) Transformador
 - D) Codificador
12. ¿Qué método se usa para transformar la variable objetivo y luego revertir la transformación?
- A) scale_target
 - B) normalize_target
 - C) inverse_transform
 - D) rescale_output
13. ¿Qué codificador crea una columna por categoría?
- A) OrdinalEncoder
 - B) FactorEncoder
 - C) OneHotEncoder
 - D) MultiColumnEncoder
14. ¿Qué interfaz se debe implementar para crear un transformador personalizado?
- A) BaseEstimator y TransformerMixin
 - B) CustomObject y PipeMixin
 - C) PipeBuilder y Estimator
 - D) TransformerABC
15. ¿Qué función evalúa un modelo con validación cruzada simple en Scikit-learn?
- A) validate_model
 - B) cross_val_score
 - C) kfold_score
 - D) model_validation

Test de Repaso - Preprocesamiento y Validación de Modelos

16. ¿Qué clase permite aplicar transformaciones diferentes por columna?

- A) ColumnProcessor
- B) ColumnHandler
- C) ColumnTransformer
- D) MultiColumnEncoder

17. ¿Qué transformador se usa para imputar valores con la mediana?

- A) KNNImputer
- B) SimpleImputer
- C) MedianFiller
- D) NullHandler

18. ¿Qué parámetro en train_test_split permite controlar la aleatoriedad?

- A) cv
- B) shuffle
- C) random_state
- D) seed

19. ¿Qué función de pandas permite discretizar variables continuas?

- A) pd.qcut
- B) pd.split
- C) pd.discretize
- D) pd.cut

20. ¿Qué técnica escala los datos a una media de 0 y desviación estándar 1?

- A) MinMaxScaler
- B) Normalizer
- C) StandardScaler
- D) RobustScaler

21. ¿Qué herramienta ayuda a evitar fugas de datos?

- A) CSV
- B) Pipeline
- C) FeatureSelector
- D) StandardScaler

22. ¿Qué parámetro define los valores a probar en GridSearchCV?

- A) param_list
- B) params
- C) param_grid
- D) grid_values

23. ¿Qué hace el scoring='neg_root_mean_squared_error'?

- A) Calcula el RMSE negativo
- B) Maximiza el error cuadrático

Test de Repaso - Preprocesamiento y Validación de Modelos

- C) Minimiza el R^2
 - D) Devuelve el MAE positivo
24. ¿Qué función de Scikit-learn permite dividir los datos en entrenamiento y prueba?
- A) `train_test_split`
 - B) `split_data`
 - C) `divide_data`
 - D) `split_train_test`
25. ¿Qué método busca todas las combinaciones de hiperparámetros?
- A) `AllSearchCV`
 - B) `FullGridCV`
 - C) `GridSearchCV`
 - D) `ExhaustiveSearch`
26. ¿Qué valor se recomienda usar para cv en validación cruzada?
- A) 2
 - B) 3
 - C) 5
 - D) 10
27. ¿Cuál es una forma común de producir data leakage?
- A) Dividir después de transformar
 - B) Transformar después de dividir
 - C) Usar `StandardScaler` dentro de pipeline
 - D) Validar sin test set
28. ¿Qué tipo de estimador es `LinearRegression`?
- A) Transformador
 - B) Clasificador
 - C) Regresor
 - D) Encoder
29. ¿Qué hace `remainder='passthrough'` en `ColumnTransformer`?
- A) Descarta columnas
 - B) Escala columnas
 - C) Pasa columnas no transformadas tal cual
 - D) Agrupa columnas por tipo
30. ¿Qué debes hacer para interpretar correctamente un RMSE negativo?
- A) Usar logaritmo
 - B) Multiplicar por 100
 - C) Sumar 1
 - D) Cambiar el signo