

Aprendizagem Automática - Pré-Processamento

1. Qual das seguintes técnicas pode ser utilizada para lidar com valores omissos num conjunto de dados?
 - a) Remover as linhas com valores em falta
 - b) Substituir os valores em falta pela mediana da coluna
 - c) Ignorar os valores em falta durante o treino do modelo
 - d) Todas as anteriores

2. No contexto de feature selection, a que se refere o termo "wrapper"?
 - a) Técnicas que avaliam a importância das características com base num critério estatístico
 - b) Técnicas que dependem do desempenho preditivo de um determinado algoritmo de machine learning
 - c) Técnicas que transformam o espaço original das features num espaço de menor dimensão
 - d) Técnicas que utilizam coeficientes de correlação para selecionar features

3. Qual das seguintes é um método de seleção de características baseado em filtros (filter-based)?
 - a) Recursive Feature Elimination (RFE)
 - b) Boruta Feature Selection
 - c) Variance Thresholding
 - d) LASSO Regression

4. Qual técnica de scaling de dados é melhor para lidar com outliers?
 - a) Min-Max Scaling
 - b) Standardização (Z-score normalization)
 - c) Max Abs Scaler
 - d) Robust Scaler

5. Qual das seguintes técnicas de scaling transforma os dados para terem uma média de 0 e um desvio padrão de 1?
 - a) Min-Max Scaling
 - b) Standardização (Z-score normalization)
 - c) Max Abs Scaler
 - d) Robust Scaler

Aprendizagem Automática não Supervisionada

1. Qual é o objetivo principal da aprendizagem automática não supervisionada?

- a) Ajustar um modelo aos dados de treino para prever resultados futuros com alta precisão.
- b) Treinar um modelo para fazer previsões com base em dados rotulados.
- c) Classificar os dados em categorias predefinidas usando exemplos de treino.
- d) Identificar padrões e estruturas nos dados sem a necessidade de rótulos ou orientação externa.

2. Em aprendizagem automática não supervisionada, o "elbow method" é frequentemente utilizado para?

- a) Calcular a distância entre pontos num espaço de alta dimensão.
- b) Determinar o número ideal de clusters num algoritmo de clustering.
- c) Calcular o número ideal de componentes num método de redução de dimensionalidade.
- d) Identificar a similaridade entre clusters em um algoritmo de clustering hierárquico.

3. Qual dos seguintes métodos de redução de dimensionalidade é linear?

- a) MDS
- b) tSNE
- c) PCA
- d) UMAP

4. Para duas execuções de clustering K-Means, é esperado obter os mesmos resultados de clustering?

- a) Sim
- b) Não

5. Quais das seguintes afirmações são verdadeiras sobre o PCA?

- 1. O PCA é um método não supervisionado de aprendizagem automática.
- 2. PCA identifica as direções que capturam a maior variância dos dados.
- 3. O número máximo de componentes principais é menor ou igual ao número de features.
- 4. Todos os componentes principais são ortogonais entre si.

- a) 1 e 2
- b) 1 e 3
- c) 2 e 3
- d) 1, 2 e 3
- e) 1, 2 e 4
- f) 1, 2, 3 e 4

6. Qual dos seguintes algoritmos é mais sensível a outliers?

- a) K-Means
- b) K-Modes
- c) K-Medoids
- d) K-Medians

7. Supõe que queres agrupar 7 observações em 3 clusters usando o algoritmo de clustering K-Means. Após a primeira iteração, os clusters C1, C2 e C3 têm as seguintes observações:

C1: {(2,2), (4,4), (6,6)}

C2: {(0,4), (4,0)}

C3: {(5,5), (9,9)}

Quais serão os centróides dos clusters se pretender prosseguir com a segunda iteração?

- a) C1: (2,2), C2: (0,0), C3: (5,5)
- b) C1: (6,6), C2: (4,4), C3: (9,9)
- c) C1: (4,4), C2: (2,2), C3: (7,7)
- d) Nenhuma das anteriores

8. O que acontece quando se diminui a dimensionalidade de um dataset usando o PCA?

- 1. As features irão manter interpretabilidade;
- 2. As features irão perder alguma interpretabilidade;
- 3. As features irão conter toda a informação dos dados;
- 4. As features irão perder alguma informação dos dados.

- a) 1 e 3
- b) 1 e 4
- c) 2 e 3
- d) 2 e 4

Aprendizagem Automática Supervisionada

1. Qual das seguintes afirmações descreve corretamente a diferença entre regressão e classificação?
 - a) Regressão prevê valores contínuos, enquanto classificação prevê categorias.
 - b) Regressão e classificação são equivalentes, mas usam algoritmos diferentes.
 - c) Classificação prevê valores contínuos, enquanto regressão prevê categorias.
 - d) Regressão usa apenas árvores de decisão, enquanto classificação usa apenas random forests.
2. No algoritmo K-Nearest Neighbors (KNN), qual dos seguintes fatores pode afetar significativamente a precisão do modelo?
 - a) O número de vizinhos escolhidos (k)
 - b) O tipo de normalização dos dados
 - c) A métrica de distância utilizada
 - d) Todas as anteriores
3. Qual das seguintes características descreve corretamente o algoritmo Naïve Bayes?
 - a) Não pode ser usado para features binárias
 - b) É baseado no Teorema de Bayes e assume independência condicional entre features
 - c) É um modelo paramétrico que sempre obtém melhor desempenho do que árvores de decisão
 - d) Não pode ser usado para classificação de textos
4. O que acontece quando se ajusta demasiadamente um modelo de regressão logística aos dados de treino?
 - a) O modelo terá alta variância e baixo viés, levando a overfitting.
 - b) O modelo terá alta precisão em dados novos.
 - c) O modelo ignorará os dados de treino e terá alta generalização.
 - d) O modelo sempre produzirá um valor de probabilidade igual a 0 ou 1.
5. Em uma árvore de decisão, qual critério pode ser usado para dividir os nós durante o treino?
 - a) Entropia e Ganho de Informação
 - b) Gini Impurity
 - c) Redução de Variância (para regressão)
 - d) Todas as anteriores
6. Nas Random Forests, por que se utilizam múltiplas árvores em vez de uma única árvore de decisão?
 - a) Para reduzir overfitting e melhorar a generalização
 - b) Para garantir que todas as árvores sejam idênticas
 - c) Para reduzir o tempo de treino
 - d) Para remover a necessidade de pré-processamento dos dados