

# Lista 1 - Aprendizado de Máquina

Thais Rodrigues

12 de dezembro de 2024

1. Existem diferentes tipos de sistemas de aprendizado de máquina. Explique as classificações a seguir e cite 1 exemplo (algoritmo) de cada.
  - a) Aprendizado supervisionado ou não supervisionado
  - b) Aprendizado baseado em instância ou baseado em modelo
  - c) Aprendizado online ou em batch
2. Comente 1 contexto no qual seria vantajoso utilizar um sistema de aprendizado de máquina e outro contexto no qual seria desvantajoso.
3. Qual é o propósito do conjunto de dados de treino, validação e teste?
4. Caso o modelo tenha um bom desempenho nos dados de treinamento, mas não nos dados de validação, o que está acontecendo? Como corrigir esse problema?
5. Qual é a diferença entre parâmetro e hiperparâmetro?
6. Como escolher o hiperparâmetro  $k$  do algoritmo KNN?
7. É necessário escalonar as covariáveis para aplicação do algoritmo KNN? Por quê? E na floresta aleatória?
8. Ajuste o KNN no conjunto de dados MNIST utilizando 2 medidas de similaridade diferentes e mensure o desempenho dos modelos.
9. Explique a medida de desempenho AUC (área sob a curva ROC).
10. Como a árvore de decisão escolhe a regra de decisão de 1 nó (covariável e valor de partição)?
11. Compare o KNN e a árvore de decisão (vantagens/desvantagens).
12. O que fazer se uma árvore de decisão estiver sobreajustando aos dados de treino?
13. Ajuste uma árvore de decisão no conjunto de dados Titanic para prever se o passageiro sobreviveu ou não ao acidente. Ajuste a profundidade da árvore por meio de um gridsearch no conjunto de validação. Avalie o desempenho do modelo.