

Curso Machine Learning
Instituto Atlântico
Avanti Bootcamp

Aluno(a): Andressa Monte

Professor: Renê Pessoa

Turma: 2024.2

Atividade 1 - Diagnóstico

1. Explique, com suas palavras, o que é machine learning?

Entendo que a machine learning (aprendizagem de máquina) diz respeito à disposição dos sistemas computacionais de se aperfeiçoarem sem uma programação específica.

O aprendizado é contínuo e é mais efetivo à medida que mais dados são incorporados e processados pelas máquinas envolvidas.

O aperfeiçoamento autônomo dos algoritmos pode ocorrer via técnicas de redes neurais, aprendizado profundo, supervisionado ou não supervisionado, aprendizado por reforço dentre outros.

Dentre os algoritmos utilizados para aprendizado, ilustram-se as árvores de decisão.

2. Explique o conceito de conjunto de treinamento, conjunto de validação e conjunto de teste em machine learning.

O conjunto de treinamento pode ser definido como um subconjunto de dados utilizados para treinar modelos de algoritmos. Tais materiais apresentam exemplos rotulados para que a máquina seja capaz de generalizar e aperfeiçoar sua performance.

Já o conjunto de validação serve para avaliar a performance do modelo algorítmico no processo de treinamento, sobretudo na etapa de generalização de dados não vistos. Ao final, esse subconjunto de dados servirá para legitimar os resultados do conjunto de teste, no qual os dados finais serão testados.

Por fim, o conjunto de teste será implementado para avaliar e processar os dados não utilizados nas etapas anteriores. Seu escopo é a performance final do modelo algorítmico, visando estimar o quão efetivo será a capacidade de generalização de dados novos pela máquina.

3. Explique como você lidaria com dados ausentes em um conjunto de dados de treinamento.

Ao encontrar os dados ausentes, buscaria primeiro remover as linhas de código que contém os valores ausentes. Se não funcionar ou não haver como excluir tais dados, tentaria imputar um volume de dados alternativo, com estimativas e dados próximos aos que foram perdidos.

4. O que é uma matriz de confusão e como ela é usada para avaliar o desempenho de um modelo preditivo?

Consiste numa tabela com resultados para avaliar a performance do modelo de classificação. Em sua estrutura estão dados de casos corretamente identificados ou não para montar comparações com os rótulos utilizados na realidade.

Os dados segmentados servirão para avaliar a acurácia do modelo preditivo, constatar erros e auxiliar correções, além de medir o nível de desempenho.

5. Em quais áreas (tais como construção civil, agricultura, saúde, manufatura, entre outras) você acha mais interessante aplicar algoritmos de machine learning?

A área da saúde exala uma necessidade imprescindível de aperfeiçoar métodos e tecnologias para melhor diagnosticar doenças, personalizar tratamentos e salvar vidas. O Machine Learning proporciona uma versatilidade de técnicas para resolver problemas complexos da área de forma rápida e personalizada.

Os modelos de Machine Learning podem ser aplicados, por exemplo, para analisar dados clínicos e imagens obtidas em radiografias e ainda monitorar a saúde de pacientes com doenças crônicas. Além disso, os algoritmos potencializam a identificação de novas moléculas para criar medicamentos e permitem o rastreamento de potenciais epidemias.

Referências: Wikipedia, Medium, livro Pattern Recognition and Machine Learning (2006)