

Tópicos Avançados em Machine Learning

A Maldição da Dimensionalidade

Quanto maior o número de atributos (dimensões), mais dados são necessários para manter a performance dos modelos. Isso pode resultar em modelos lentos e com baixo desempenho. Problemas como overfitting se tornam mais comuns.

Soluções:

- Seleção de features
- Redução de dimensionalidade

Engenharia e Seleção de Features

Etapa essencial para melhorar o desempenho e a interpretação do modelo.

- Seleção: Escolher variáveis mais relevantes com base em critérios estatísticos ou algoritmos.
- Transformação: Modificar features existentes (normalização, codificação, etc.).
- Criação: Gerar novas variáveis com base nas existentes (combinações, decomposições, etc.).
- Extração: Reduzir a dimensionalidade mantendo a informação essencial (ex: PCA).

Underfitting vs Overfitting

- Underfitting: Modelo muito simples, não aprende bem os dados de treino -> baixa acurácia.
- Overfitting: Modelo muito complexo, aprende demais os dados de treino -> baixa generalização.

Objetivo: Encontrar o ponto ideal entre os dois extremos.

Viés vs Variância

- Viés: Erro causado por suposições incorretas no modelo (baixa complexidade).
- Variância: Sensibilidade do modelo às variações nos dados (alta complexidade).

Tópicos Avançados em Machine Learning

O ideal é balancear viés e variância para alcançar um modelo generalizável.

Ensembles de Modelos

Combinação de vários modelos para melhorar a performance geral:

- Bagging: Usa subconjuntos dos dados para treinar modelos independentes e combina as previsões.
- Boosting: Treina modelos sequencialmente corrigindo erros anteriores.
- Stacking: Usa previsões de modelos como entrada para um modelo final.
- Voting: Usa votação para decidir a saída final.

Estrutura de Projetos em IA/ML

- CRISP-DM: Processo iterativo com 6 etapas (entendimento do negócio, dados, preparação, modelagem, avaliação, deploy).
- ML Canvas: Ferramenta visual para organizar e comunicar o projeto entre técnicos e não técnicos.
- AI Canvas: Abordagem mais centrada no impacto humano e estratégico da IA.