



EXERCÍCIO 11

SISTEMAS DE RECOMENDAÇÃO

Objetivo

Neste exercício, faremos um experimento completo dentro do contexto de Sistemas de Recomendação.

Sistemas de Recomendação são ferramentas muito presente em nossas vidas digitais. É muito difícil utilizarmos qualquer serviço web ou acessarmos sites e não visualizarmos nenhuma forma de recomendação, como filmes para assistir e produtos para comprar.

Dos diferentes tipos de recomendação, a Filtragem Colaborativa é uma das mais usadas atualmente em aplicações acadêmicas e comerciais. Isso ocorre devido a sua facilidade de aplicação, visto que depende apenas de uma matriz de interações usuário-item para gerar a recomendação, sem demandar dados adicionais (como é na recomendação baseada em conteúdo).

Ao término deste exercício, espera-se que você entenda como um experimento completo na área de sistemas de recomendação pode ser conduzido, com enfoque em filtragem colaborativa. Você irá implementar dois métodos tradicionais da área, KNN e SVD, e avaliá-los de diferentes maneiras, mensurando sua capacidade de prever uma nota e gerar uma lista top-N.

O exercício

Ao longo do exercício, você deverá completar 5 funções:

- `knn`: responsável pelo cálculo de similaridade entre itens, execução do KNN e predição de nota;
- `svd_sgd_optimizer`: responsável por ajustar um par de vetores de fatores latentes de um usuário e um item, utilizando a técnica de *stochastic gradient descent*;
- `rmse_mae`: responsável por calcular o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE) de uma dada predição;
- `precision_recall_f1`: responsável por calcular a Precisão, a Revocação e a F-Medida de um dado ranqueamento top-N;

- **ndcg**: responsável por calcular o Ganho Cumulativo Descontado Normalizado (NDCG) de um dado ranqueamento top-N.

Preencha o código apenas nos espaços delimitados por comentários, normalmente iniciados por um comentário “COMPLETE O CÓDIGO AQUI” e instruções para a implementação.

As implementações devem ser genéricas e funcionar para qualquer conjunto de dados. Na avaliação, as funções serão testadas em bases com quantidade de usuários e itens diferentes das fornecidas com o exercício, tendo em comum apenas as dimensões e nomes das colunas (no caso de funções que recebem DataFrame como entrada).

Os casos de teste

Este exercício possui **5 casos de teste** que buscam avaliar cada uma das funções implementadas. A distribuição de tarefas é feita da seguinte forma:

- **Caso de teste 1:** corrige a função `knn`;
- **Caso de teste 2:** corrige a função `svd_sgd_optimizer`;
- **Caso de teste 3:** corrige a função `rmse_mae`;
- **Caso de teste 4:** corrige a função `precision_recall_f1`;
- **Caso de teste 5:** corrige a função `ndcg`.