

Bootcamp: Arquiteto(a) de Machine Learning

Desafio Prático

Módulo 3: Sistemas de Recomendação

Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Recomendação por filtragem colaborativa.
2. Avaliação de modelos.

Enunciado

Em um ambiente de desenvolvimento em Python (recomenda-se o Google Colaboratory), implemente as funções descritas para responder às questões objetivas relativas à seguinte função utilidade:

Figura 1 – Função utilidade para 5 usuários e 7 itens, com avaliações codificadas no intervalo real de 1 a 5. Os valores nulos representam avaliações desconhecidas.

	I1	I2	I3	I4	I5	I6	I7
U1	4.0	0.0	0.0	4.7	1.0	0.0	0.0
U2	5.0	4.5	4.0	0.0	0.0	0.0	0.0
U3	0.0	0.0	0.0	1.5	5.0	4.0	0.0
U4	4.1	3.0	0.0	4.9	0.0	0.0	3.0
U5	1.0	4.0	0.0	2.5	3.8	1.0	5.0

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Acessar o ambiente [Google Colaboratory](#) (recomendado) ou qualquer ambiente de desenvolvimento em Python.
2. Implementar as seguintes funções em Python:

Figura 2 – Função para cálculo da similaridade por cossenos.

```
def cosine_similarity(x: np.array, y: np.array):  
    """Calcula a similaridade de cossenos entre os arrays x e y."""  
  
    cosine_sim = np.dot(x, y)/(np.linalg.norm(x) * np.linalg.norm(y))  
  
    return cosine_sim
```

Figura 3 – Funções para cálculo da similaridade por cossenos centralizada.

```
def array_centering(v: np.array):  
    """Subtraindo dos elementos não nulos pela média."""  
    # cópia para evitar sobrescrita  
    v = v.copy()  
  
    # indexação para extrair elementos não nulos  
    non_zeros = v > 0  
  
    # substituição pela média  
    v[non_zeros] = v[non_zeros] - np.mean(v[non_zeros]) + 1e-6  
  
    return v  
  
def centered_cosine_similarity(x: np.array, y: np.array):  
    """Calcula a similaridade de cossenos centralizada entre os arrays x e y."""  
  
    # subtraindo a média  
    x = array_centering(x)  
    y = array_centering(y)  
  
    # similaridade por cossenos  
    centered_cosine_sim = cosine_similarity(x, y)  
  
    return centered_cosine_sim
```

3. Responder às seguintes questões objetivas:

- a. Qual o valor da similaridade de cossenos entre os usuários U1 e U2?
- b. Qual o valor da similaridade de cossenos entre os usuários U1 e U3?
- c. Qual o valor da similaridade de cossenos entre os usuários U1 e U4?
- d. Qual o valor da similaridade de cossenos entre os usuários U1 e U5?
- e. Qual o valor da similaridade de cossenos centralizada entre os usuários U1 e U2?
- f. Qual o valor da similaridade de cossenos centralizada entre os usuários U1 e U3?
- g. Qual o valor da similaridade de cossenos centralizada entre os usuários U1 e U4?
- h. Qual o valor da similaridade de cossenos centralizada entre os usuários U1 e U5?
- i. Considerando a similaridade de cossenos centralizada, qual o usuário é mais similar ao usuário U1?
- j. Considerando a similaridade de cossenos centralizada, qual o segundo usuário mais similar ao usuário U1?
- k. Considerando uma filtragem colaborativa User-User, com agregação pela média simples e número de vizinhos igual a 2, qual a predição para a avaliação do usuário U1 ao item I2?
- l. Considerando uma filtragem colaborativa User-User, com agregação pela média simples e número de vizinhos igual a 2, qual a predição para a avaliação do usuário U1 ao item I1? Suponha que a avaliação (U1, I1) seja desconhecida.

- m. Qual o erro absoluto ($|\text{real} - \text{previsto}|$) da predição da tupla (U1, I1), considerando uma filtragem colaborativa User-User, com agregação pela média simples e número de vizinhos igual a 2?
- n. Considere um sistema de recomendação cuja predição r é dada por $r = b_u + b_i - \mu$, onde b_u é a média simples das avaliações conhecidas do usuário u , b_i é a média simples das avaliações conhecidas do item i , μ é a média simples de todas as avaliações conhecidas. Quais os valores de b_u , b_i e μ para o usuário U1 e item I2?
- o. Considere um sistema de recomendação cuja predição r é dada por $r = b_u + b_i - \mu$, onde b_u é a média simples das avaliações conhecidas do usuário u , b_i é a média simples das avaliações conhecidas do item i , μ é a média simples de todas as avaliações conhecidas. Considerando este modelo, qual a predição para a avaliação do usuário U1 ao item I2?