

AVALIAÇÃO #1

1. **(3.0)** Em um recomendador colaborativo baseado em avaliações explícitas, qual o propósito de normalizar uma avaliação r_{ui} pela média das avaliações históricas do usuário u ? E pelo desvio padrão dessas avaliações? Justifique.
2. **(3.0)** Um usuário assiste a 10 seg de um vídeo de 10 min e navega para outra página. Como essa observação pode ser explorada como feedback para a modelagem das preferências desse usuário? Justifique.
3. **(3.0)** Considere um usuário u e dois itens i e j distintos **não** consumidos por u . Qual desses itens mais provavelmente representa um exemplo negativo das preferências de u ? Justifique.

Para as questões a seguir, considere as matrizes (a) e (b) e fórmulas de referência (c) abaixo.

	i_0	i_1	i_2	i_3	i_4	i_5
u_0		4		3		4
u_1			4		5	3
u_2	1	5			3	
u_3	5	4	5			
u_4	2			5		2
u_5	2		5			4
u_6			5		1	1
u_7			5	2	3	
u_8	4			2	3	
u_9		4	5	1		

(a) avaliações de treino

	i_0	i_1	i_2	i_3	i_4	i_5
u_0	4		3		5	
u_1	5	1		1		
u_2			4	1		4
u_3				2	2	3
u_4		5	1		2	
u_5		1		4	1	
u_6	5	1		3		
u_7	2	2				2
u_8		5	3			2
u_9	2				4	4

(b) avaliações de teste

Predição baseada em item para o usuário u e o item i :

$$\hat{r}_{ui} = \frac{\sum_{j \in \mathcal{N}_{ui}} s(\vec{i}, \vec{j}) r_{uj}}{\sum_{j \in \mathcal{N}_{ui}} |s(\vec{i}, \vec{j})|}$$

- \mathcal{N}_{ui} : itens mais próximos a i avaliados por u
- $s(\vec{i}, \vec{j})$: similaridade entre os itens i e j
- r_{uj} : avaliação do usuário u sobre o item j

Similaridade (cosseno) entre os vetores \vec{i} e \vec{j} :

$$s(\vec{i}, \vec{j}) = \frac{\sum_{p=1}^d i_p j_p}{\sqrt{\sum_{p=1}^d i_p^2} \sqrt{\sum_{p=1}^d j_p^2}}$$

- i_p : p -ésima dimensão do vetor \vec{i}
- j_p : p -ésima dimensão do vetor \vec{j}
- d : número de dimensões dos vetores

Root mean squared error (RMSE) para o usuário u :

$$\text{RMSE}_u = \sqrt{\frac{1}{n} \sum_{p=1}^n (r_{up} - \hat{r}_{up})^2}$$

- r_{up} : avaliação do usuário u sobre o p -ésimo item
- \hat{r}_{up} : predição para o usuário u e o p -ésimo item
- n : número de itens no ranking

Discounted cumulative gain (DCG) para o usuário u :

$$\text{DCG}_u = \sum_{p=1}^n \frac{2^{r_{up}} - 1}{\log_2(p+1)}$$

- r_{up} : avaliação do usuário u sobre o p -ésimo item
- n : número de itens no ranking

(c) fórmulas de referência

4. **(6.0)** Para um dado usuário-alvo u_x , onde x é o último dígito de seu número de matrícula (e.g., se seu número de matrícula é 2021081017, seu usuário-alvo é u_7), **recomende itens previamente não-avaliados por u_x usando o algoritmo de filtragem colaborativa baseado em item, a similaridade do cosseno, e uma vizinhança de tamanho $k = 2$** . Para cada item recomendado, indique claramente (1) a predição computada e (2) os cálculos intermediários que levaram a essa predição. A matriz de avaliações mostrada acima **não** deve ser normalizada.
5. **(5.0)** Calcule os valores de RMSE e DCG para as recomendações produzidas na questão anterior para o usuário u_x . Os cálculos intermediários também deverão ser apresentados.

