

AVALIAÇÃO #1

1. **(1.5)** Um usuário assiste a 15 seg de um vídeo de 13 min de duração e navega para outra página. Como esse evento pode ser explorado como feedback em um sistema de recomendação baseado em filtragem colaborativa?
2. **(1.5)** Por que a avaliação média recebida por um item é um critério inapropriado para recomendação não-personalizada?
3. **(1.5)** Considere os seguintes cenários quanto ao número de usuários (m) e itens (n) de um sistema de recomendação:
 - Cenário 1: $m = 10$, $n = 1.000.000$
 - Cenário 2: $m = 1.000.000$, $n = 10$
 - Cenário 3: $m = 1.000.000$, $n = 1.000.000$

Que tipo de recomendação colaborativa (user-based ou item-based) seria a mais indicada em cada cenário? Justifique.

4. **(1.5)** Na separação das avaliações disponíveis em treino e teste, quais as vantagens e desvantagens de utilizar as avaliações mais recentes (em vez de avaliações aleatoriamente selecionadas) para teste?
5. **(1.5)** Por que as tarefas de predição e ranking são tipicamente avaliadas utilizando métricas diferentes?

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		3
u_1			1		3	4
u_2	1	4			2	
u_3	2	3	3			
u_4	4			4		2
u_5	5		1			5
u_6			1		2	4
u_7			5	1	3	
u_8	2			1	4	
u_9		3	5	4		

(a) avaliações de treino

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

(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

6. **(4.5)** 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 2021081014, seu usuário-alvo é u_4), **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 $k = 3$** . 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.
7. **(3.0)** 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.