

# 1 Problem 1 (20 pts)

## Theoretical Task

In the coupled factorization, assume you have item features  $Y$  of dimensionality  $n_y < d$ , where  $d$  is the optimal rank of decomposition.

Which representation of coupled factorization form will be more economic in terms of memory usage for the item cold-start scenario:

- when mapping is applied to the latent item features;

$$\mathcal{L}_{coupled}(A, \Theta) = \|A - PQ^\top\|_F^2 + \|X - PG^\top\|_F^2 + \|Y - QW^\top\|_F^2$$

- when mapping is applied to the initial item features?

$$\mathcal{L}_{coupled}(A, \Theta) = \|A - PQ^\top\|_F^2 + \|X - PG^\top\|_F^2 + \|YW - Q\|_F^2$$

Where  $X \in \mathbb{R}^{M \times m_x}$  is the matrix of user attributes;

$Y \in \mathbb{R}^{N \times n_y}$  is the matrix of item features;

$\Theta = \{P, Q, G, W\}$ ;

$P \in \mathbb{R}^{M \times d}$  is the matrix of latent user features;

$Q \in \mathbb{R}^{N \times d}$  is the matrix of latent item features;

$G \in \mathbb{R}^{m_x \times d}$  is the mapping from user features to user latent features;

$W \in \mathbb{R}^{n_y \times d}$  is the mapping from item features to item latent features.

The first representation [1] requires less memory usage compared to the second representation [2] when mapping to latent item features. This is because the  $Y$  matrix in the first representation has a dimensionality of  $Nn_y$  and  $QW^\top$  is of dimensionality  $N \times n_y$ , resulting in the need to store  $N \times n_y$  values. On the other hand, the second representation [2] involves matrices  $W$  and  $Y$  of dimensionality  $n_y d$  and  $Nn_y$ , respectively. When these matrices are multiplied, the resulting product  $YW$  has a dimensionality of  $N \times d$ . The matrices  $Q$  and  $YW$  together requires more memory  $N \times d$ . So more memory required for the second representation to store these parameters. Also to find optimal  $q^*$  we need to solve equation  $qW^\top = y^*$  in [1]. In [2] we need to solve  $q = yW$ . In second case we need to transpose, so again, second case need more memory.

As a result, when using coupled factorization and mapping to latent item features [1], the representation will have a more efficient memory usage in the item cold-start scenario.