

Problem Set 13, Dec 10, 2020 (Solutions to Matrix Factorization Theory Questions)

Problem 3 (Which models will give the best performance):

Intuitively, the Movie Mean baseline should perform the best amongst all the baselines because it should be able to tell whether a movie is liked or disliked by the audience in general.

Problem 4 (Implement Stochastic Gradient Descent):

NOTE: We use MSE instead of RMSE for updates.

1. Let $E = X - WZ^\top$. Then, $e_{dn} = x_{dn} - (WZ^\top)_{dn}$.

Then, the entry corresponding to the i^{th} row of W , w_i in $\nabla_{(W,Z)} f(W, Z)$ is given by $\frac{1}{|\Omega|} \sum_{(i,n) \in \Omega} e_{in} z_n$. Similarly, the entry corresponding to z_j in $\nabla_{(W,Z)} f(W, Z)$ is given by $\frac{1}{|\Omega|} \sum_{(d,j) \in \Omega} e_{dj} w_d$.

2. Given a single rating x_{dn} , G will be a sparse matrix. The entry corresponding to W_d will be $\frac{1}{|\Omega|} e_{dn} z_n$ and that corresponding to Z_n will be $\frac{1}{|\Omega|} e_{dn} w_d$. Rest of the entries will be zero.

7. One can observe that only those entries of WZ^\top need to be updated whose row-index corresponds to an updated row of W or column-index corresponds to an updated row of Z . Thus, in order to avoid the complete matrix computation after every step, we only need to recompute the dot product of only those row pairs of W and Z in which at least one amongst the pair has been updated and update the corresponding entries of WZ^\top .

Problem 5 (Alternating Least Squares):

The entry corresponding to w_i in $\nabla_{(W)} L(W, Z)_{ij}$ is given by $\sum_{(i,n) \in \Omega} e_{in} z_n$. Setting it to zero and solving, we get $\sum_{(i,n) \in \Omega} x_{in} z_n = \sum_{(i,n) \in \Omega} (w_i^\top z_n) z_n$. Now, note that $(w_i^\top z_n) z_n$ can be re-written as $(z_n z_n^\top) w_i$. Using this in the previous equation, we get $w_i = \sum_{(i,n) \in \Omega} (z_n z_n^\top)^{-1} \sum_{(i,n) \in \Omega} x_{in} z_n$.

Similarly, we obtain $z_j = \sum_{(d,j) \in \Omega} (w_d w_d^\top)^{-1} \sum_{(d,j) \in \Omega} x_{dj} w_d$.

In case of regularized ALS, the updates for w_i and z_j become $\sum_{(i,n) \in \Omega} (z_n z_n^\top + \lambda I_k)^{-1} \sum_{(i,n) \in \Omega} x_{in} z_n$ and $\sum_{(d,j) \in \Omega} (w_d w_d^\top + \lambda I_k)^{-1} \sum_{(d,j) \in \Omega} x_{dj} w_d$ respectively where k is the number of latent features.