Labs
Machine Learning Course
Fall 2020

EPFL
School of Computer and Communication Sciences
Nicolas Flammarion & Martin Jaggi
www.epfl.ch/labs/mlo/machine-learning-cs-433

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 $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{W}\boldsymbol{Z}^\top$. Then, $e_{dn} = x_{dn} - (\boldsymbol{W}\boldsymbol{Z}^\top)_{dn}$.

Then, the entry corresponding to the i^{th} row of $\boldsymbol{W}$, w_i in $\nabla_{(\boldsymbol{W},\boldsymbol{Z})} f(\boldsymbol{W},\boldsymbol{Z})$ is given by $\frac{1}{|\Omega|}\sum_{(i,n)\in\Omega} e_{in} z_n$. Similarly, the entry corresponding to z_j in $\nabla_{(\boldsymbol{W},\boldsymbol{Z})} f(\boldsymbol{W},\boldsymbol{Z})$ 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 $\boldsymbol{W}\boldsymbol{Z}^\top$ need to be updated whose row-index corresponds to an updated row of $\boldsymbol{W}$ or column-index corresponds to an updated row of $\boldsymbol{Z}$. 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 $\boldsymbol{W}$ and $\boldsymbol{Z}$ in which at least one amongst the pair has been updated and update the corresponding entries of $\boldsymbol{W}\boldsymbol{Z}^\top$.

Problem 5 (Alternating Least Squares):

The entry corresponding to w_i in $\nabla_{(\boldsymbol{W})} L(\boldsymbol{W},\boldsymbol{Z})_{ij}$ 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 $(\sum_{(i,n)\in\Omega} z_n z_n^\top + \lambda \boldsymbol{I}_k)^{-1} \sum_{(i,n)\in\Omega} x_{in} z_n$ and $(\sum_{(d,j)\in\Omega} w_d w_d^\top + \lambda \boldsymbol{I}_k)^{-1} \sum_{(d,j)\in\Omega} x_{dj} w_d$ respectively where k is the number of latent features.