

Assignment 6: Recommendation Systems

October 18, 2022

Submit your assignment in python 3.6 or above.

For this assignment, you will use the movielens data set from here:
<https://grouplens.org/datasets/movielens/latest/>.

There are 2 data sets, small and large. Use both datasets.

Do the following for each dataset:

1. Apply SVD on each of these data sets. How many latent factors do you need to capture most of this information? Do the latent factors correspond to some intuition that you have?
2. Apply CUR decomposition to these datasets. For a given number of latent factors, what is the additional error introduced by CUR decomposition as opposed to SVD?
3. What is the running time and storage required by CUR decomposition as opposed to SVD?
4. Plot the error versus number of latent factors.
5. Plot the storage and compute time.
6. Apply PQ decomposition or the machine learning approach to matrix factorization. Do a 80:20 split of your data into training and test data. Apply stochastic gradient descent from scratch with regularization. Play around with the regularization hyperparameter. What kind of training error did you get with PQ decomposition.
7. Try the generalized neural collaborative filtering approach. What is the architecture of your neural network? Apply a 2 layer multi layer perceptron as your NCF architecture. What is the dimension of the neural embedding for users and movies versus the number of dimensions you picked for matrix factorization?

Submit you code as recommender-student-name.py. Write a report and submit as recommender-student-name.pdf