Machine Learning

Programming Assignment 2

Recommender System using Matrix Factorization

Name: Yaswanth Janardhan Konduru

ID: 1001741845

1. Your approach to use the matrix factorization algorithm as a recommender system.

I have used matrix factorization algorithm to build this recommender system. We have a 2m dataset but due to the limitation of resources I had to reduce the dataset to 10lakhs, to improve the performances and reduce the runtimes of the algorithm. In matrix factorization we take 2 random matrix and the actual data matrix and then using the algorithm we try to predict the null values in the data matrix. Which are the predictions these values are calculated with an epoch of 100 where each epoch we calculate the error value and the try to reduce the error in each epoch which will help us in get the most accurate values. Once this is done, we pass the user id which will get the predicted values of the user.

1. What is the maximum dataset that your recommender system can use?

Here I have ran the dataset of size 10lakhs, but my recommender system can run up to a dataset of 50lakh (It also depends on the resources available) data which ran for almost 24 hours for 100 epochs. So, I limited the dataset to 10 lakhs data and epoch 100 these are the runtimes.

100%|██████████| 100/100 [45:02<00:00, 27.02s/it]

For 100 epochs

1. What is the time complexity of your recommender system?

The time complexity of matrix factorization is O(n3).

1. What is the performance of your recommender system?

The time taken for the recommender system to run the matrix factorization with following hyper parameters

* Epoch : 100
* Features: 3
* Alpha: 0.002
* Beta: 0.02

100%|██████████| 100/100 [45:02<00:00, 27.02s/it]

For 100 epochs

1. Is there a way to scale-up your recommender system to work with very large datasets?

Resources available, Evaluation tools play a very important role in processing large amount of data as 2 million dataset is very huge for normal computers to process. We can try different methods like content based methods, collaborative filtering methods (model based, memory based), hybrid methods(mix of multiple methods) that are more optimized to use for recommender systems.

Sample Outputs:

Initial matrix from the dataset -

Table

Description automatically generated

Matrix after predicting -

Graphical user interface, application, table

Description automatically generated

Sample output with used id - 25

Graphical user interface, application

Description automatically generated

References :

https://medium.com/sfu-cspmp/recommendation-systems-collaborative-filtering-using-matrix-factorization-simplified-2118f4ef2cd3

https://towardsdatascience.com/recommender-systems-matrix-factorization-using-pytorch-bd52f46aa199

https://analyticsindiamag.com/how-to-build-your-first-recommender-system-using-python-movielens-dataset/

https://beckernick.github.io/matrix-factorization-recommender/

https://albertauyeung.github.io/2017/04/23/python-matrix-factorization.html

https://towardsdatascience.com/understanding-matrix-factorization-for-recommender-systems-4d3c5e67f2c9

https://towardsdatascience.com/recommendation-system-matrix-factorization-d61978660b4b