

CSC311Project

April 1, 2024

A

3.Matrix Factorization

(a)

I ran SVD with $k = 1, 2, 5, 10, 20$ and calculated the validation accuracy for each k . The best k was 5 with an accuracy of 0.659046006209427.

Validation Accuracy: 0.6428168219023427 with $k = 1$

Validation Accuracy: 0.6579170194750211 with $k = 2$

Validation Accuracy: 0.659046006209427 with $k = 5$

Validation Accuracy: 0.6586226361840248 with $k = 10$

Validation Accuracy: 0.6539655659046006 with $k = 20$

Best k : 5 with test accuracy: 0.6635619531470506

(b)

One limitation of SVD for this task is that it treats missing entries by using mean imputation. However, mean imputation can introduce noise, especially in sparse regions of the data, as it assumes that missing values are similar to the mean of the respective item. This might not hold true in reality, especially for different users or items. Therefore, a limitation of SVD in this case is that it may lead to incorrect estimations of missing values, thereby affecting the quality of matrix reconstruction.

(c d)

After implementing the ALS with SGD, I ran the algorithm with $k = 1, 2, 5, 10, 20$ and learning rate = 0.01, number of iterations = 10.

The best k was 20 with an accuracy of 0.7070279424216765.

Validation Accuracy: 0.70575783234547 with $k = 1$

Validation Accuracy: 0.7054755856618685 with $k = 20$

Validation Accuracy: 0.7046288456110641 with $k = 50$

Validation Accuracy: 0.7067456957380751 with $k = 100$

Validation Accuracy: 0.7061812023708721 with $k = 200$

Best k: 100 with Test Accuracy: 0.7067456957380751

(e)



Figure 1: ALS with SGD

B

3.Matrix Factorization