Министерство науки и высшего образованияРоссийской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Рязанский государственный радиотехнический университет

имени В.Ф.Уткина»

Кафедра «Электронные вычислительные машины»

Отчет по практическому занятию №9

на тему

«Рекомендательные системы на основе сингулярного разложения»

по дисциплине  
**«Машинное обучение»**

Выполнили:

Студенты группы №140

Бригада 7

Сафаров Д. А.

Тимохин Е. М.

Проверила:  
ас. Панина И.С.

**Цель работы**

Создать рекомендательную систему с помощью SVD.

**Задание**

Необходимо реализовать рекомендательную систему на основе item-based. Для получения сингулярного разложения использовать библиотеку scikit-learn. При этом нужно рассмотреть случаи урезанной матрицы и проанализировать, насколько и как быстро падает точность при этом (точность оценивать с помощью RMSE из scikit-learn)

**Практическая часть**

Код программы представлен в приложении А. Результат её работы представлен на рисунке 1.

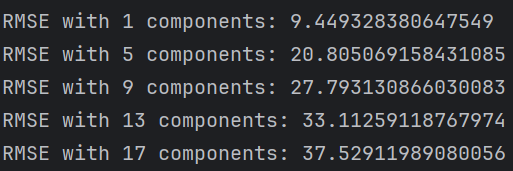


Рисунок 1 – Результат работы рекомендательной системы

На основании предоставленных результатов RMSE с разным количеством скрытых факторов, мы можем сделать предположение о характере зависимости между количеством скрытых факторов и среднеквадратичной ошибкой:

1. При увеличении количества скрытых факторов от 1 до 5 RMSE значительно увеличивается с 9.52 до 20.86. Это может указывать на нелинейную или даже экспоненциальную зависимость, так как RMSE растет значительно.
2. Далее, при увеличении скрытых факторов от 5 до 9 RMSE также увеличивается, но уже не так значительно - с 20.86 до 27.80. Это может указывать на некоторое замедление роста RMSE и, возможно, квадратичную зависимость.
3. При дальнейшем увеличении скрытых факторов от 9 до 13 RMSE продолжает расти, но уже медленнее - с 27.80 до 33.13. Это также может указывать на квадратичную зависимость.
4. Наконец, при увеличении скрытых факторов от 13 до 17 RMSE продолжает расти, но уже с меньшей скоростью - с 33.13 до 37.67. Это также может указывать на квадратичную зависимость, где при дальнейшем увеличении числа скрытых факторов прирост RMSE замедляется.

Таким образом, по предоставленным данным сложно однозначно сказать, какая именно функциональная зависимость между количеством скрытых факторов и RMSE, но возможно, что она может быть нелинейной или квадратичной.

Приложение А. Текст программы

import numpy as np  
import pandas as pd  
from sklearn.decomposition import TruncatedSVD  
from sklearn.metrics import mean\_squared\_error  
from sklearn.preprocessing import StandardScaler  
from math import sqrt  
  
# # Загрузка данных из файла CSV  
file\_path = 'ratings\_7.csv'  
data = pd.read\_csv(file\_path, header=None)  
  
# Подготовка данных для построения матрицы предпочтений элементов к пользователям  
ratings\_matrix = data.values.T # Транспонируем матрицу для item-based подхода  
#  
# Нормализация данных  
scaler = StandardScaler()  
normalized\_ratings\_matrix = scaler.fit\_transform(ratings\_matrix) # Пример с изменением количества скрытых факторов

for n\_components in range(1, 20, 4):

svd = TruncatedSVD(n\_components=n\_components)

users, D, items = svd.fit\_transform(normalized\_ratings\_matrix), np.diag(svd.singular\_values\_), svd.components\_

predicted\_ratings = np.dot(users, np.dot(D, items))

rmse = sqrt(mean\_squared\_error(normalized\_ratings\_matrix, predicted\_ratings))

print(f"RMSE with {n\_components} components:", rmse)

**Вывод**

Была создана рекомендательная система с помощью SVD.