

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

path = r"C:\Users\alepa\Desktop\Data Science\[1] 2nd Semester\Mathematical Models and Numerical Methods for Big Data\Code\Movie_Rating.csv"
ratings_df = pd.read_csv(path)

# some data cleaning after looking at the spreadsheet
# drop label columns, replace -1 with NaN since we will center data
# and -1s would be affected and lose their purpose
ratings_only_df = ratings_df.drop(['Risposta', 'Nome'], axis=1).replace(-1, np.nan)

# center the data by subtracting the mean rating for each movie
centered_ratings_df = ratings_only_df.subtract(ratings_only_df.mean(), axis=1)

# plug -1s back after centering
centered_ratings_df.fillna(-1, inplace=True)

# define num_features as 2 as we need a 2D graph
num_features = 2
A = centered_ratings_df

# Funk SVD algorithm
def funk_svd_qr(A, num_features, steps=5000, learning_rate=0.0002, beta=0.02):
    num_rows, num_cols = A.shape
    Q = np.random.rand(num_rows, num_features)
    R = np.random.rand(num_features, num_cols)
    for step in range(steps):
        for i in range(num_rows):
            for j in range(num_cols):
                if A.iloc[i, j] != -1: # this is so that our re-plugged -1
values don't get in the way
                    pred_ij = np.dot(Q[i, :], R[:, j])
                    e_ij = A.iloc[i, j] - pred_ij
                    Q[i, :] += learning_rate * (2 * e_ij * R[:, j] - beta *
Q[i, :])
                    R[:, j] += learning_rate * (2 * e_ij * Q[i, :] - beta *
R[:, j])
    return Q, R

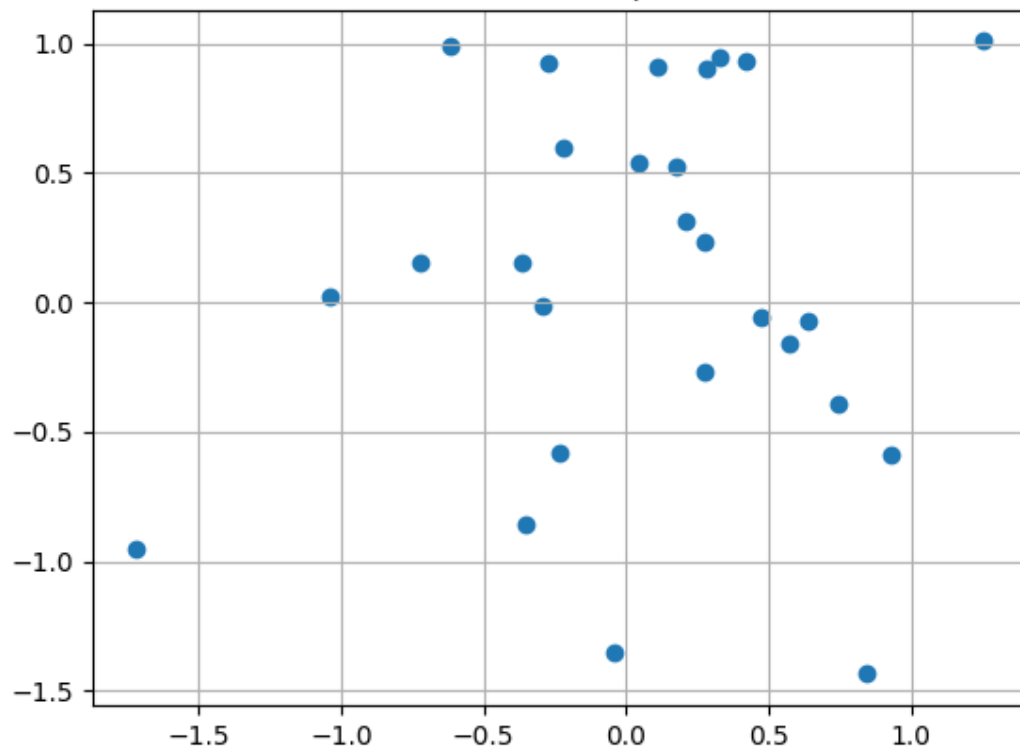
# execute the function
Q, R = funk_svd_qr(A, num_features)

# plotting
plt.scatter(Q[:, 0], Q[:, 1])
plt.title('User feature space')
plt.grid(True)
plt.show()

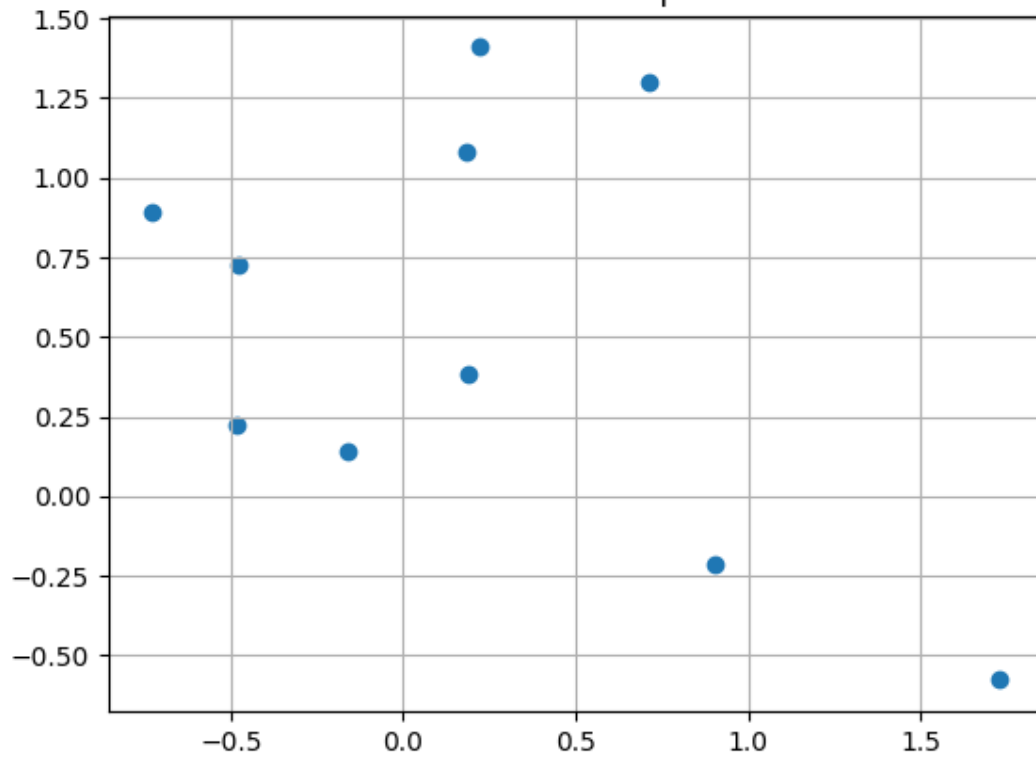
plt.scatter(R[0, :], R[1, :])
plt.title('Movie feature space')
plt.grid(True)
plt.show()

```

User feature space



Movie feature space



From these visualizations we cannot extract a strong correlation as far as the Movie feature space is concerned; we may notice three small clusters in the top left of the graph, but interpretations on movie 'quality' similarity would be a stretch.

As far as the User feature space is concerned, however, we see a stronger general trend, and we may say that, on average, users tended to rate movies not entirely sparsely but similarly.