Go Back to Recommendation Systems

Course Content

Singular Value Decomposition (SVD)

In the previous pre-read, we understood collaborative filtering, where we used similarities between users or items to recommend items to users. However, with a highly sparse user-item interaction matrix, where the majority of entries are missing or 0, the collaborative filtering strategy might produce unsatisfactory results.

There is another method to perform collaborative filtering using matrix factorization where we can identify the relationship between items and users for sparse matrices as well. With the input of users' ratings on the items, we can predict how the users would rate the items. This way the users can get the recommendation based on the ratings.

The idea behind matrix factorization is to represent users and items in a lower-dimensional space and **extract hidden features** from the data which are constructed by some hidden relations. These hidden features are called **Latent features**.

The latent features cannot be observed but can be extracted using matrix factorization algorithms. One of the most common and useful matrix factorization algorithms is **Singular Value Decomposition.**

Singular Value Decomposition:

In SVD, we decompose a matrix into 3 small matrices that are relevant to the original matrix. These matrices can be used to reconstruct the original matrix. Usually, the user-item interaction matrix has many missing values, so the purpose of SVD is to estimate the matrix values by filling null values with 0.

SVD decomposes the original matrix X into the following form:

$$X = USV^T$$

Where,

X = The original user-item interaction matrix with size $m \times n$.

U = Matrix of Latent Features for Users with size $m \times r$.

V^T = Matrix of Latent Features for Items with size $r \times n$.

S = It is a diagonal matrix of singular values with size $r \times r$.

The shape of U and V must be $m \times r$ and $r \times n$, respectively, because for matrix multiplication the number of columns of the first matrix must be equal to the number of rows of the second matrix.

Now, let’s consider an example of movies rated by users. Suppose the user-item interaction matrix looks like the below figure:

USERS	Star Trek	Alien	Serenit y	Titanic	Amelie
USER 1	1	1	1	0	0
USER 2	3	3	3	0	0
USER 3	4	4	4	0	0
USER 4	5	5	5	0	0
USER 5	0	2	0	4	4
USER 6	0	0	0	5	5
USER 7	0	1	0	2	2

Here, movies StarTrek, Allien, and Serenit belong to the Sci-Fi genre while Titanic and Amelie belong to the romantic genre. We can see that User 1, User 2, User 3, and User 4 have only rated Sci-Fi movies out of which User 4 has given the highest rating. Similarly, User 5, User 6, and User 7 have given comparatively higher ratings to romantic movies, out of which User 6 has the highest ratings for romantic movies.

Suppose, we got the following matrix U of latent features for the users:

USERS	F1	F2	F3
USER 1	**0.13**	-0.02	-0.01
USER 2	**0.41**	-0.07	-0.03
USER 3	**0.55**	-0.09	-0.04
USER 4	**0.68**	-0.11	-0.05
USER 5	0.15	**0.59**	**0.65**
USER 6	0.07	**0.73**	**-0.67**
USER 7	0.07	**0.29**	**0.32**

We can see that the values in the 1st column for User 1, User 2, User 3, and User 4 are 0.13, 0.41, 0.55, and 0.68, respectively and the values for these users in the other 2 columns are comparatively lower. We know that these users show interest in Sci-Fi movies. Hence, the information of the original matrix is captured.

Similarly, we can see that the values in the 1st column for User 5, User 6, and User 7, are 0.15, 0.07, and 0.07, respectively, and the values for these users in the other 2 columns are comparatively higher. We know that these users show interest in Romantic movies. Hence, the information of the original matrix is captured.

And similarly, suppose, we got the following matrix V^T for latent features of movies:

	Star Trek	Alien	Serenit y	Titanic	Amelie
F1	**0.56**	**0.59**	**0.56**	0.09	0.09
F2	-0.12	0.02	-0.12	**0.69**	**0.69**
F3	**0.40**	**-0.80**	**0.40**	0.09	0.09

We can see that the value in the 1st row and the 3rd row for movies Star Trek, Alien, and Serenity are comparatively higher than the 2nd row. We know that these movies are Sci-Fi while the other 2 movies are romantic.

Similarly, we can see that the values in the 2nd row are comparatively higher for Titanic and Amelie than in the other 3 rows. We know that these movies are of the Romantic genre. Hence, the information is being captured from the original matrix.

Note: In reality, we have a large number of users and items, which makes the latent features non-interpretable.

And the sigma matrix S is given as,

12.4	0	0
0	9.5	0
0	0	1.3

Here, each value in the diagonal represents the i^{th} singular value which indicates the strength of the i^{th} latent feature.

With these 3 matrices, we can reconstruct and approximate the original matrix by using the formula $X = USV^T$.

Truncated SVD

Truncated SVD produces a factorization where the number of columns can be specified for several truncations. For example, given an $n \times n$ matrix, truncated SVD generates the matrices with the specified number of columns, whereas SVD outputs n columns of matrices. The truncated SVD better works on the sparse matrices for feature output.

$$\widehat{L}_{ij} = \frac{1}{\hat{p}} \sum_{k=1}^{r} s_k u_{ik} v_{jk}, \quad \text{for all } i, j$$