IITM-CS5691 : Pattern Recognition and Machine Learning
New Method Writeup

Release Date: 29th Apr, 2021
Due Date: 28th May, 2021

Roll No: CS18B015

Name: G SHYAM SUNDAR

- Dear Student, You may have tried or thought of trying different methods for your data contest. Choose one of the methods that was not taught in class, and submit a writeup of this "new method" in the template provided below. This is an individual submission - i.e., while you would've done your kaggle submission as a team of two members or while you may've discussed this method with your teammate, **you will have to write about the new method in your own words independently and submit it individually**.
- **Template:** Fill in whatever fields are applicable for your algorithm (overall 1-2 page writeup; since some fields may not be applicable for certain methods, we haven't shown points below).

1. (points) [Name of Method, and its ML Problem and Paradigm: problem could be regression/classification/clustering/etc., and paradigm could be supervised/unsupervised/..., generative/discriminative/direct, linear/non-linear models, etc.)]:

   Solution:
   Name Of the Method: Item - Item Collaborative Filtering.
   It is regression based problem and paradigm is Supervised.

2. (points) [Brief introduction/motivation: One paragraph to describe briefly the new method (its name, what it does, its main application, etc.)]

   Solution:
   Item - Item Collaborative Filtering is normally used for recomendation systems. It is based on finding the similarties between the items(songs in the data contest) by the ratings given by the user. So when we want to predict a rating a user gives to Item. We find the top N items the user has rated and do a weigted similarity average (Which will be discussed testing algo section) to find the rating.

3. (points) [Closely related method seen in class, and relation of your selected new method to method you eventually used for the data contest]:

Solution:
This method is similar to what we have to do in data contest Here The Users are the customers and the Items are the songs But, To Use the metadata Like the features of the Song I treated these Features as Users Like for example consider the Number of comments, I will scale the value to 1 and 5. and treat It User to find the similarities between songs.

4. (points) [Training Input and Output: (e.g., $\{x_i, y_i\}_{\{i=1...N\}}, x_i \in R^d, y_i \in \{-1, +1\}$, etc.]:

Solution:
Let N be the number of Users and M be the number of Items.

Training Input : User - Item utility Matrix ($N \times M$).

Let $A \epsilon \mathbb{R}^{N \times M}$ be the User - Item utility Matrix then the value of A_{ij} will be rating given by $User_i$ to $Item_j$.

Trainig Output : Similarity Matrix ($M \times M$).

Let $S \epsilon \mathbb{R}^{M \times M}$ be the similarity Matrix then the value of S_{ij} will be the Similarity between $Item_i$ and $Item_j$

So The diagonal values will be naturally 1.

5. (points) [Training Objective function (e.g., loss function that is to be optimized) or probabilistic model (over which MLE or Bayesian inference done):]

Solution:
No Objective Function.

6. (points) [Training Algorithm: Brief description of key aspects of the algorithm]

Solution:

The main aspect of the Training Algorithm is finding the similarity Matrix.

We calculate the cosine similarity between the Items based on User ratings given to it.

As some users May be highly opmistic that 3 often may be there least rating and some are highly pessimistic that the same 3 may be there highest rating so we normalise the data by subtracting the average user rating for items which the user has rated.

Let v_i and v_j be the vectors of size N (Number of users).

v_i corresponds to the ith column in User - Item Similarity Matrix But normalized as said above. and All the Null values (i.e the Users who have not rated $Item_i$) will be made to zero. similarly with v_j

Now we calculate the cosine similarity which is cos of the angle between the two Item vectors v_i and v_j .

This is how we calculate Similarity between two Items. And we do It all the pairs and find the Similarity Matrix.

7. (points) [Testing Input and Output: (e.g., $x \in R^d$, $y \in \{-1, +1\}$)]

Solution:
Testing Input : The ID's of User and the Item for which the rating is to be predicted.

Testing Output : Rating between (1 - 5).

8. (points) [Testing Algorithm: Brief description of key aspects of the algorithm]

Solution:
We predict the rating by doing the weigted similarity average.

So At first we for the given User Id we find the Items which he rated.

And take the simlarities with those Items with the Item which we need to predict the rating from Similarity Matrix we computed in the Training Part.

So, Now the weighted Average would be $\frac{\sum_{i=1}^{n} S_i R_i}{\sum_{i=1}^{n} S_i}$

where S_i is the similarity and R_i is the rating.

This would be the predicted rating.

9. (points) [Critique of the method: (1-2 paragraphs discussing its strengths and weaknesses in your own words)]

Solution:

This Algorithm fails when a new user comes in as we go by the previous ratings given by the User. But when a new Item comes in we can predict.