PRML DATA CONTEST

D.Mithun CS16B004
Muqeeth EE16B026

April 26, 2019

Models Used

- Base line model
- Collaborative filtering using neighbourhood model
- SVR Regression on movie features

Explanation on models

- Firstly implemented base line model , A baseline estimate for an unknown rating r_{ui} is denoted by b_{ui} where b_{ui} is predicted as $b_{ui} = b_u + b_i + \mu$. where $b_u$and b_i are the observed deviations of user u and item i from the average. Implemented this using batch gradient descent method by going around 20 iterations (epochs) and conidering the value of η as 0.01 and batch size of 128.
- Next improved this baseline model using neighbourhood model where we used similarities between movies and predict the unknown rating r_{ui}which is denoted by $\hat{r}_{ui}$as $\hat{r}_{ui} = b_{ui} + ((\Sigma_{j \epsilon S^k(i;u)} s_{ij}(r_{uj} - b_{uj}))/(\Sigma_{j \epsilon S^k(i;u)} s_{ij}))$ where $s_{ij} = (n_{ij} / (n_{ij} + \lambda)) * \rho_{ij}$, where s_{ij}measures the similarity between movies i and j and $\rho_{ij}$is the correlation coefficient between movie i and movie j and typical value for λ is taken as 100 and n_{ij} is number of users who rated both movie i and j.
- Then used svr regression on movie features, To predict the rating given by a particular user u on movie i if a user u rated atleast one movie then considering the movie features of all the movies rated by user u and trained the model using svr regression after applying pca on movie features, if user u has not rated any movie then applied svr regression on all movie features given and predicted the average rating by considering rating of that movie as average rating of all users who rated that movie.
- movies with movie features values are not given are taken as mean of all given movie features.
- Finally rating is predicted by taking the average of neighbourhood model and regression model.

Model tried but not used:

- Movie genres are used as features for regression on the movies. It didnot show the improvement in mse. So we didn't used this model.
- We tried to implement the latent factor model but the values are not converging.
- Initially we used linear regression on movies, later we changed to svr regression which improved our model.

MSE values

Below are the mean squared error values we got for the different models we used

Model	Mean square error
Base line Model	0.83-0.84
Neighbourhood Model	0.82-0.83
Average of Neighbourhood Model and Regression Model	0.73-0.74