PRML 2019 Data Contest Report

Predict Movie Ratings

CS5691 : Pattern Recognition and Machine Learning (JAN-MAY 2019)

Course Instructor: Harish Guruprasad Ramaswamy

INDIAN INSTITUTE OF TECHNOLOGY

MADRAS, CHENNAI

TEAM 1

Gaurav Singhal (cs18s032)

Rahul Biswas(cs18s008)

Introduction:

A recommender system refers to a system that is capable of predicting the future preference of a set of items for a user, and recommend the top items. In our case it's Predict the ratings of movies by different users.

Approaches:

- Mean Rating Prediction:

  In this approach we just return the constant value for all the predictions, which is the global mean of all previously rated movies. We were getting the Mean square error on test set 0.97862

- Baseline Prediction:

  In this approach we have implemented baseline model. We have calculated a_u's for each user and b_i's for each item present in train dataset. Using these values we calculated the values for test set. We were getting MSE on test data 0.82636 - 0.83096

  - Hyper param settings:
    - Epocs = 10
    - Mini batch size = 500
    - Eta = 0.005
    - Reg param = 0.02

- Baseline Prediction with Regression and PCA :

  In this approach we have used above baseline model for predicting the rating of items for which we have some information about previous ratings in train data and Regression for the remaining items using to genome data available. Using PCA we reduce dimension to 200 and then apply regression

  Different regression models used are:

  - Linear Regressor
  - Random Forest Regressor
  - Decision Tree Regressor
  - Support Vector Regressor

- Neighbourhood Prediction Model :

  We have tried to implement neighbourhood model but to create correlation matrix it was taking large amount of time so we didn't use this approach in our prediction model.

- Latent Factor Prediction With Regression and PCA :

  In this approach we have implemented Latent factor model using Single Value Decomposition (SVD) without considering available genome data available for movies. We have merged this model with Regression of PCA extracted attribute set. We have used this model as our FINAL submission with approx MSE of 0.78692 .

  - Hyperparameters
    - Hidden factors size = 70
    - Epocs = 20
    - Reg param = 0.02
    - Eta = 0.005
  - PCA settings

- Reduced dimension 200

- Latent Factor(with genome):

  In this approach we have implemented Latent factor model using Single Value Decomposition (SVD) considering available genome data available for movies. But because of some bugs and time constraint we were not able time implement it.