

258 Assignment 1 Report, Name: Ting-Chou Lin, PID: A53317103

- Kaggle User Information:
 - Display Name: JamesTcl
 - User Name: k2973363
 - Email Address: til002@eng.ucsd.edu
- Read prediction
 - Ranking:
 - Public Leaderboard: 404/847
 - Private Leaderboard: 387/847
 - Details of approaches for submitted results:
 - Ensemble learning with logistic regression
 - $Features = [book\ popularity, Jaccard\ book\ similarity, Jaccard\ user\ similarity]$
 - $Popularity_{book} = (\# Read_{book} - minimum\ \# Read) / (maximum\ \# Read - minimum\ \# Read)$
 - $maximum\ \# Read$ and $minimum\ \# Read$ are constants over the entire dataset
 - For the similarities, instead of only counting the maximum one as the feature, I use the mean of top K similarities as the feature
 - Parameters of submitted results:
 - $\# of\ randomly\ choosed\ negative\ samples\ in\ training\ set = 100168$
 - Logistic Regression regularization strength, $C = 0.01$
 - Top K of counted similarites, $K = 5$
 - Things that don't help:
 - Feature scaling or Cosine similarities as features
 - Logistic regression with `class_weight = 'balanced'`
 - Using the mean of all the similarities but not the top K ones
 - Support vector machine (SVM) or K-nearest neighbors (KNN)
 - Using the normalized number of reading book per user as a feature
- Rating prediction
 - Ranking:
 - Public Leaderboard: 42/423
 - Private Leaderboard: 39/423
 - Details of approaches for submitted results:
 - Simple (bias only) latent factor-based recommender
 - $rating(user, item) \approx \alpha + \beta_{user} + \beta_{book}$
 - The optimization problem
 - $argmin \sum_{u,b} (\alpha + \beta_{user} + \beta_{book} - R_{u,b})^2 + \lambda_{user} \times \beta_{user}^2 + \lambda_{book} \times \beta_{book}^2$
 - Separate different lambdas for regularizing the two bias terms
 - Parameters of submitted results:
 - $\lambda_{user} = 0.00001$
 - $\lambda_{book} = 0.0001$
 - Things that don't help:
 - Use linear regressor with features of Jaccard similarities and book popularities