London Business School

AM11 2022-23

Final Assessment

Part 3

The objective of part 3 is to build some models based on Sessions 6-10 that can predict movie rating using genre and features generated in Part 2 of the assignment. In the Part 3 folder you can find the features (*features_fromPart2.RDS*) that have been generated from Part 2 in the case that you were not able to fully complete Part 2.

Consider the set of movies that Part 2 is focusing on. From the https://grouplens.org/datasets/movielens/25m/ work with the "ratings.csv" file calculate the average rating a movie gets. Also from the same folder download the "movies.csv" file that contains genre information for each movie. Create 20 dummy variables (0 or 1) that indicate whether the movie belongs to the specific genre or not.

After this pre-processing you should have an average rating per movie and 20 (genre dummies) + 10 (PCAs from Part 2). Use this dataset to complete the following tasks.

First task:

Train an SVM model that predicts whether a movie gets an excellent rating (i.e., average rating over 3.75/5). Perform Hyperparameter tuning (cost parameter C and kernel hyperparameter) to better the performance of your model. Present the hyperpameters of your best possible model. Discuss problems of underfitting or overfitting of your model (if any). Compare your findings with another classifier that you are familiar with (eg., logistic regression). How much the performance of your SVM increases/decreases if instead of 3.75 you use 4 as the threshold of an "excellent movie"? Discuss.

Second task:

Train an ANN (Artificial Neural Network) that predicts average rating per movie (regression). Perform Hyperparameter tuning* to increase the performance of your model. Present the architecture and hyperpameters of your best possible model. Compare your findings with another linear regression. Briefly discuss your choices.

[*] Choose only two hyperparemeters to do your search.