Algo we live by - assignment2

March 2018

Description

For this assignment, I used three models Logistic regression, Multiple neural network and Decision tree to train given data.
Step 1: By using "importdata()" function, I cleaned and imported features X, and label Y. To be mentioned, I used several methods to transform 4 features. As result, this function could provide nb1, nb2, r1, r2, log(nb1), log(nb2), nb1-nb2, r1-r2, log(nb1/nb2) as features for future training and prediction.
Step 2: By using "feature_selection" function, I could transform X to most related feature sets for prediction according to different models.
Step 3: By using "cross_validation" function, it will play cross_validation training with 4 folders on data set, and return accuracy for each dataset prediction.
Conclusion:
For Logistic regression model, if without feature transformation, but with normalize function, the average accuracy for each folder is roughly 63%. However, if use log(nb1) and log(nb2) to replace nb1 and nb2 respectively, the average accuracy could increase to roughly 73%.
For Decision tree model, feature transformation doesn't influence result much. It ends up with accuracy from 64% to 67% using different transformation of features.
For neural network classification model, inspired by paper professor recommended, I used 2 layers and each layer with 50 neural nets. It achieved with accuracy from 70% to 73% according to different feature transformation.