Tutorial 7

In this tutorial you will learn how to take advantage of ensemble learning in order to improve the performance of regression tasks. By the end of this tutorial you will be able to train various ensemble models and use them for prediction.

Part 1: Regression

1- Load the training and test datasets and randomForest library.
2- Train a randomForest model over the training data using 10 trees.
3- Calculate the error rate for the randomForest predictor over the test set.
4- Next repeat steps 2 and 3 using 100 trees. Observe whether performance improves.
5- In the second half, train a multi-variate glm regression model and an SVM model over the training set and predict over the test set. Calculate the error rate for the two predictors.
6- Create an ensemble model by combining the predictions of SVM and glm regression equally. Calculate the error rate of this ensemble model.
7- Repeat step 6, except in this case, we put a heavier weight to the prediction of SVM.

Hand in: In ensembling SVM with glm, how do error rates change when the weights of the predictions of the two models change? RandomForest is itself an ensemble model. Can RandomForest be added to the ensemble of SVM and glm?