Machine Learning Crash Course

Task

Regression

1. Check that you can use `ML_demo.m` to obtain root-mean-square test error for linear regression on a 75%/25% train/test split of the Portuguese wine dataset.
2. By using the `Treebagger` function, repeat the same experiment, replacing linear regression with random forest regression. *Hint: make sure you choose the regression option rather than classification.*
3. Investigate whether the random forest prediction can be improved by changing the minimum leaf size.
4. Improve your code by averaging over 10 randomly chosen 75%/25% train/test splits.

Classification

5. Read the Pima Indians diabetes database, `diabetes.csv`, into Matlab. *The classification task is to build a model which predicts whether patients have diabetes based upon the other eight diagnostic measurements.*
6. Using the testing framework that you have developed, test the k-nearest neighbours classifier on the diabetes classification task. Replace root-mean-square error with proportion of correct classifications.
7. Investigate the effect of (a) changing the number of neighbours and (b) standardizing (normalizing) the predictors.
8. Try out random forest and logistic regression classifiers on the same problem. *Hint: use* `glmfit` *and* `glmval` *for logistic regression.*
9. *Taking things further*: using the deep learning toolbox or otherwise, investigate using a multi-layer perceptron network with a single hidden layer on each of the two tasks.