

Supervised Machine Learning Analysis

My Name, My Email

December 19, 2014

Introduction

Data acquisition

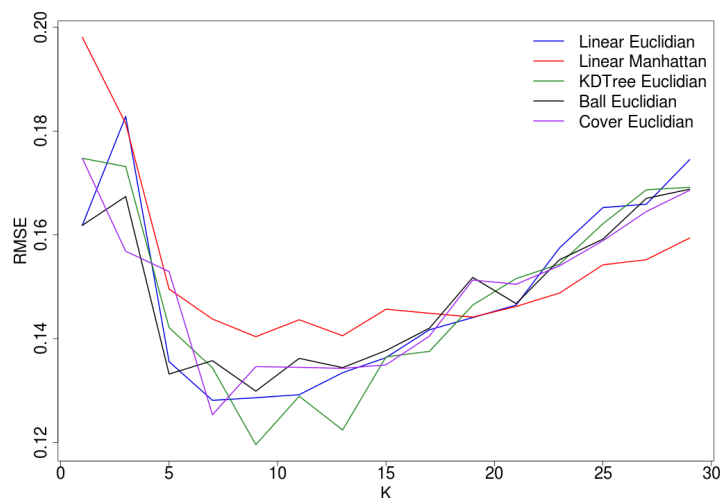
Dataset

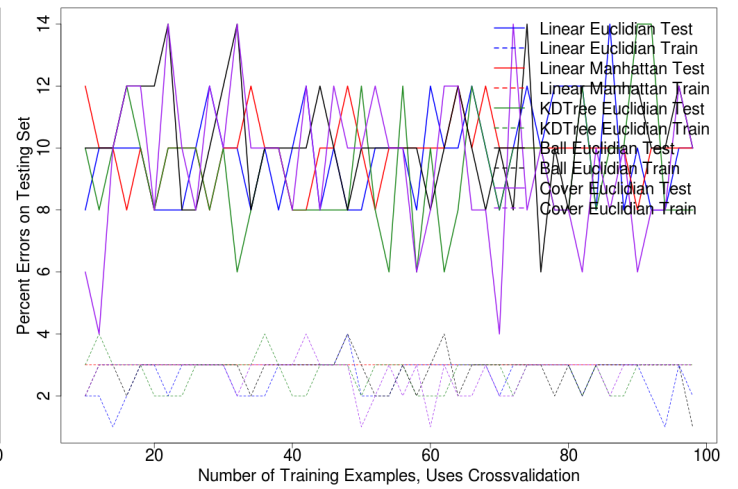
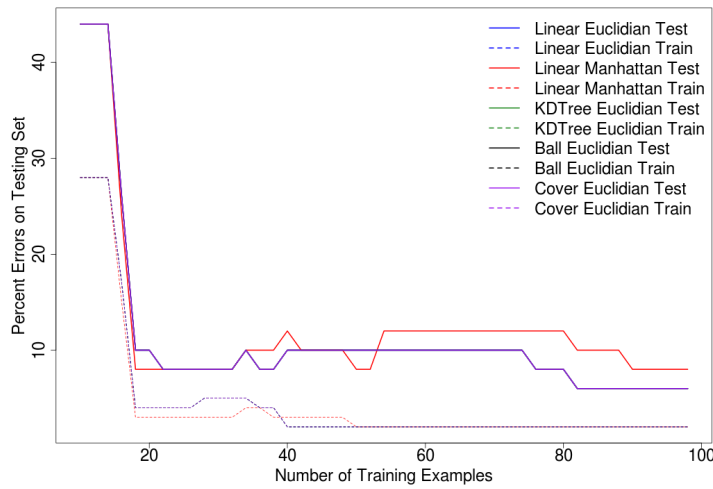
Tools Used and Methodology

A jython interface to Weka[?] is used to find error using different amounts of training data against a fixed test set. This data is plotted using R and finally a latex report is generated.

K-Nearest Neighbors

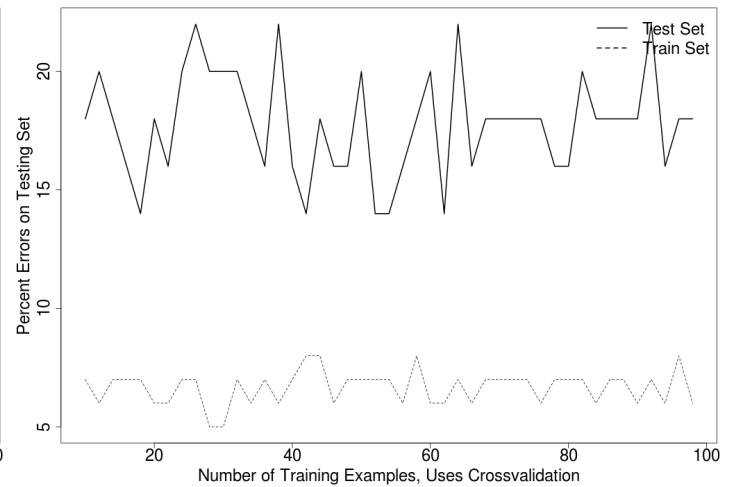
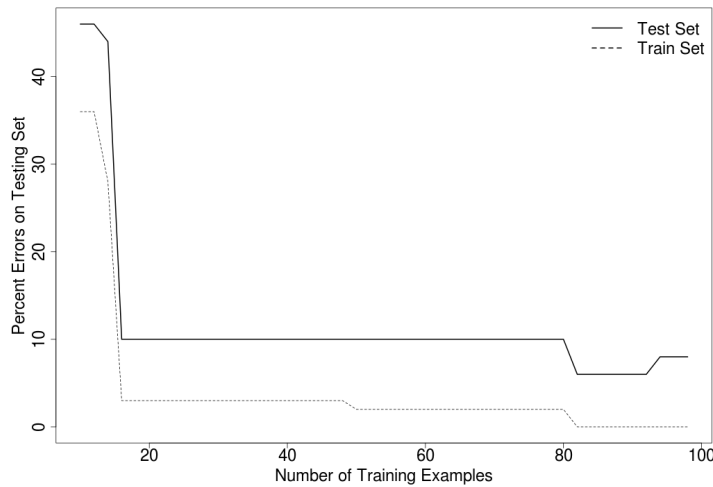
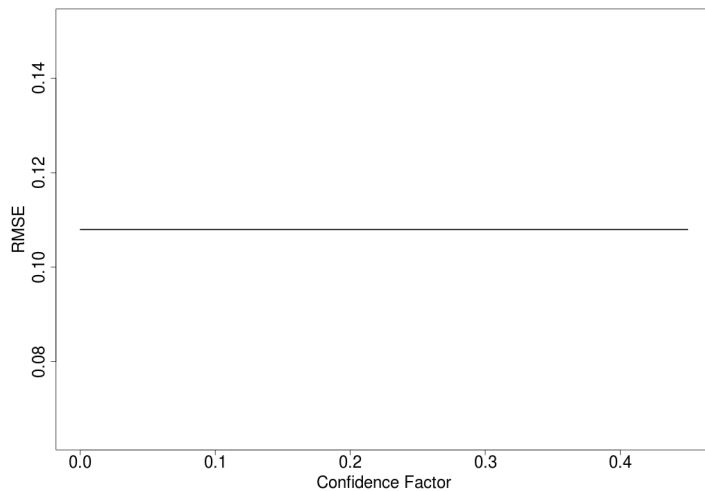
Find an optimal K by looking at the plot below. Plug this value of K into ml.properties file. Once domain knowledge values have been set for all algorithms, re-run the data generation.





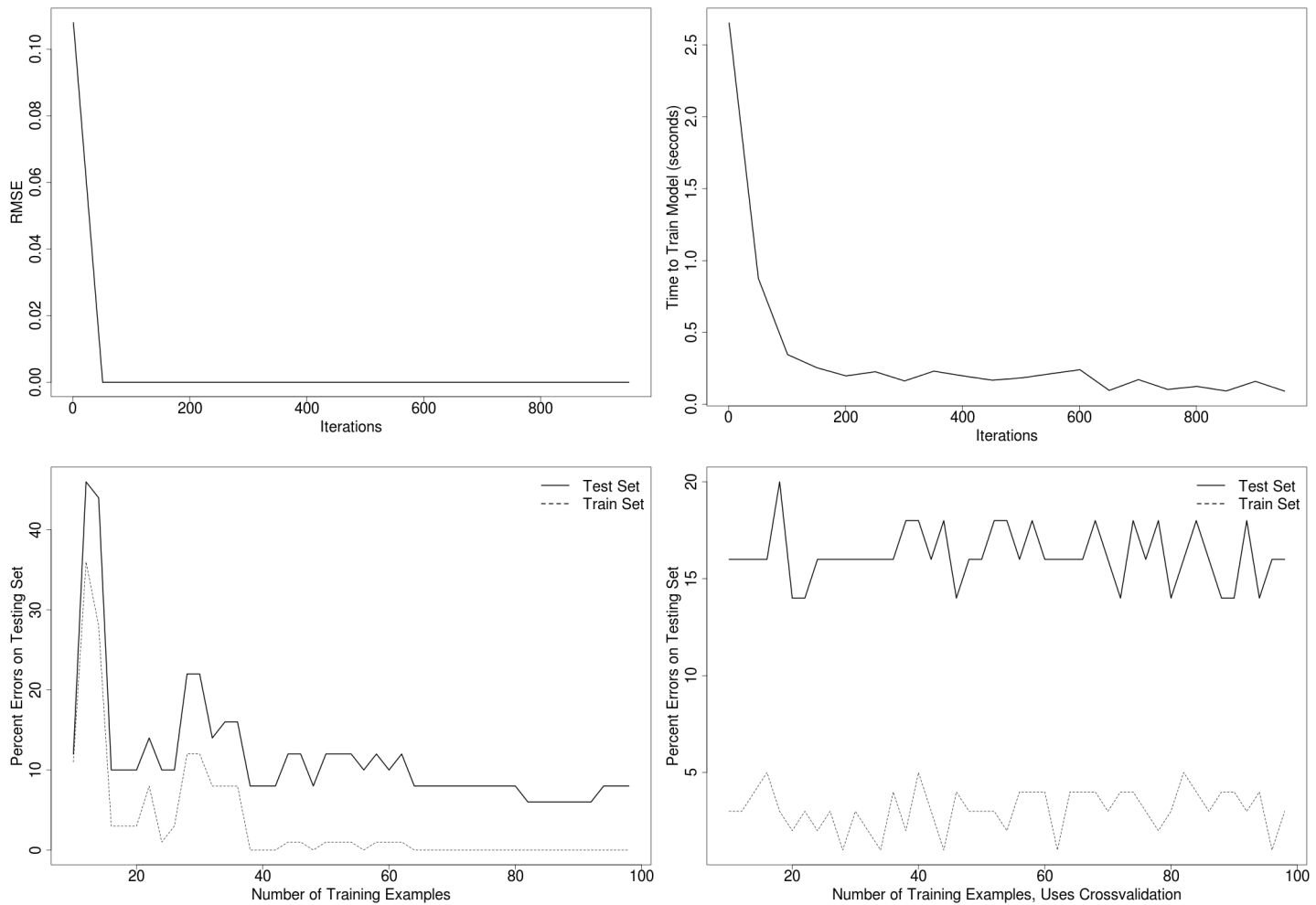
Decision trees with pruning

Find an optimal confidence factor by looking at the below plot. Plug this value into ml.properties file. Once domain knowledge values have been set for all algorithms, re-run the data generation.



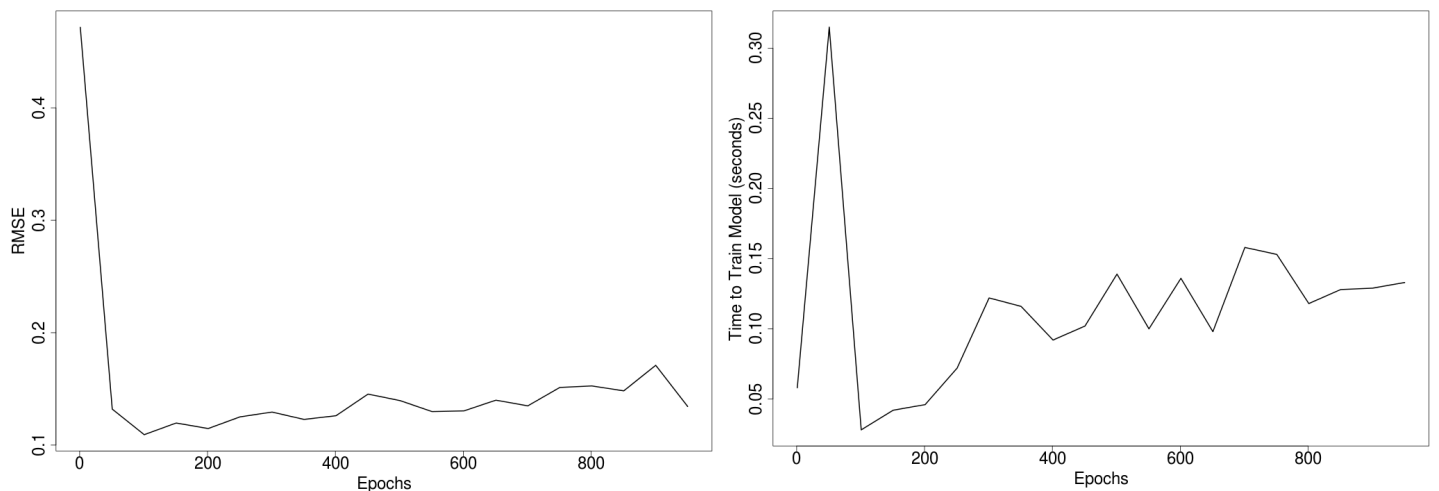
Boosting

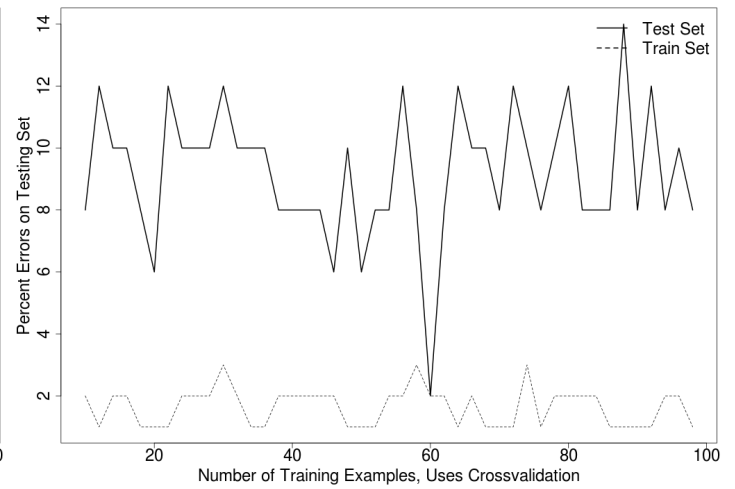
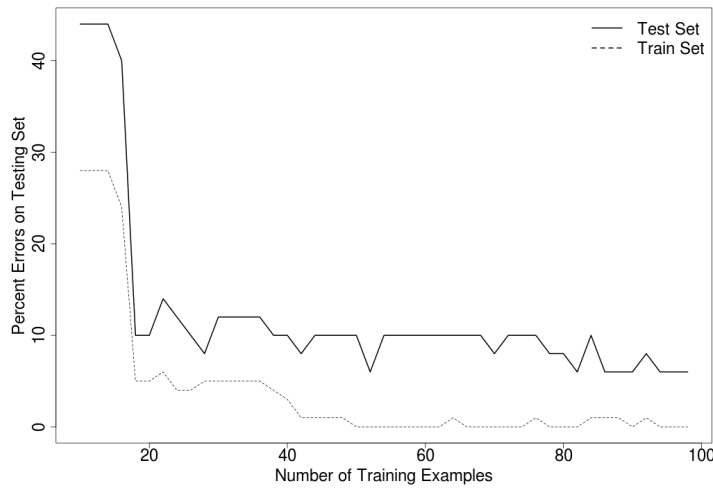
Find an optimal value for number of iterations and then plug it into ml.properties file. Once domain knowledge values have been set for all algorithms, re-run the data generation.



Neural Networks

Find an optimal value for number of epochs and set it in ml.properties file. This can help avoid long run times. Once domain knowledge values have been set for all algorithms, re-run the data generation.





Support Vector Machines

You need to find an optimal value of C and then plug it into the `ml.properties` file. Once domain knowledge values have been set for all algorithms, re-run the data generation.

