# Exploring Classification with Weka

In this assignment we used the machine learning exploratory tool Weka to compare methods for classi-fication. The dataset Satimage was used, which contains six different classes. We found that SVM had the lowest misclassification error of any of the methods (with the parameter values shown in the table), while still maintaining a fast computation speed. Below is a table of misclassification errrors for train and test sets, as well as runtime for each algorithm. Also included is a plot of test errors vs. computation speed.

Method	Training Set		Testing Set	
	Minimum error	Runtime (s)	Minimum error	Runtime (s)
Decision tree	0.024	0.25	0.148	0.25
Random Forest (100 trees)	0	3.46	0.086	3.10
Random Forest (300 trees)	0	9.75	0.087	9.75
Logistic Regression	0.119	2.84	0.162	2.84
Multilayer Perceptron	0.048	35.30	0.125	36.22
Naïve Bayes	0.203	0.02	0.204	0.02
Adaboost (20 trees)	0	5.39	0.102	5.75
Adaboost (100 trees)	0	27.62	0.084	29.24
LogitBoost (10 stumps)	0.122	1.20	0.153	1.07
LogitBoost (100 stumps)	0.034	11.71	0.118	11.55
LogitBoost (100 stumps, 0.95 trim)	0.033	3.58	0.124	3.96
LogitBoost (25 regression trees)	0	117.78	0.083	105.35
SVM (c = 3, gamma = 5)	0.033	3.85	0.078	4.01

Misclassification Error vs. Runtime
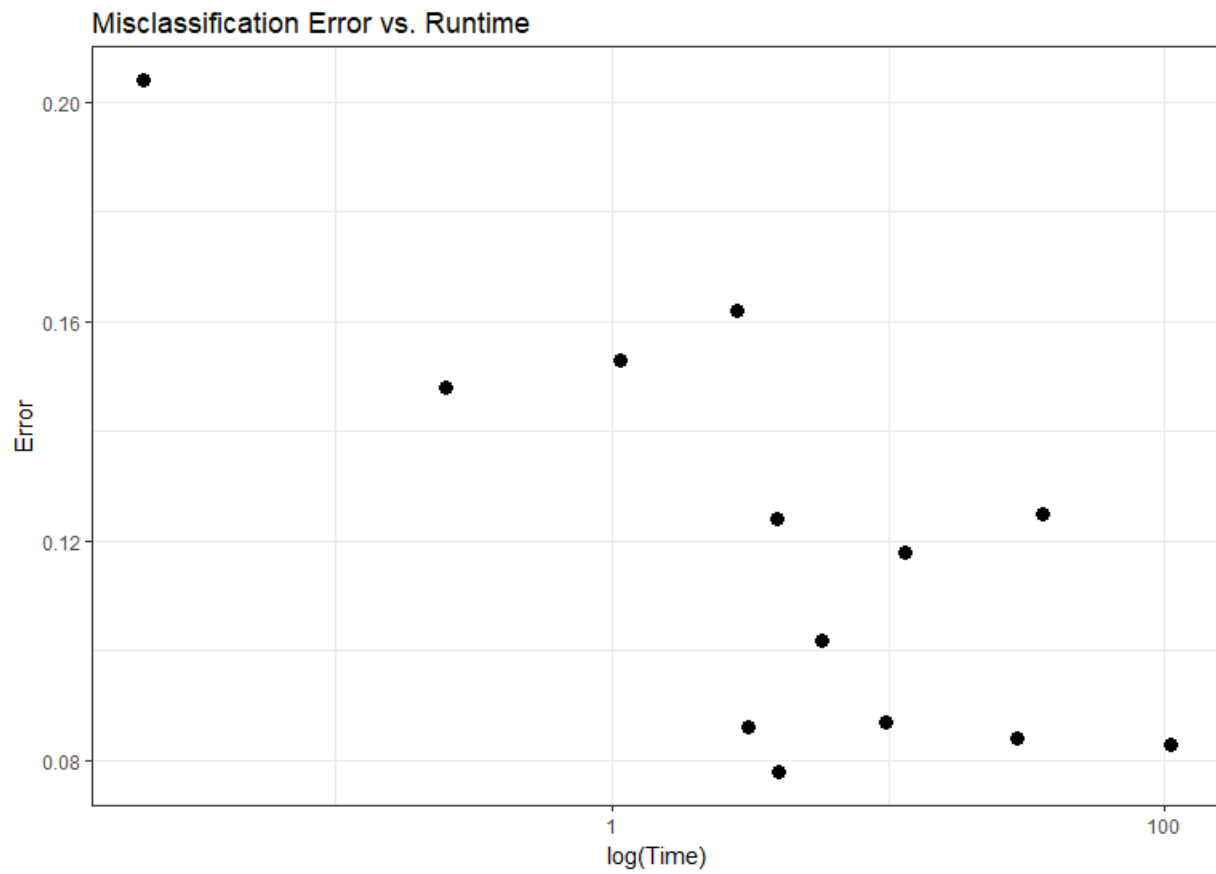
Bibliography

Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.