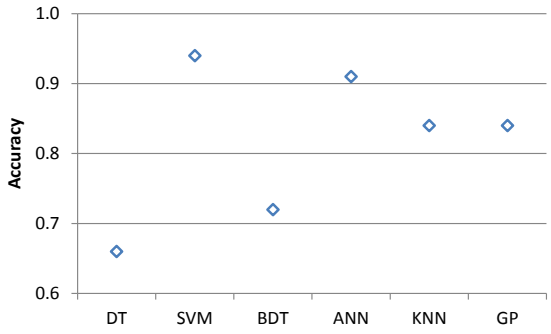
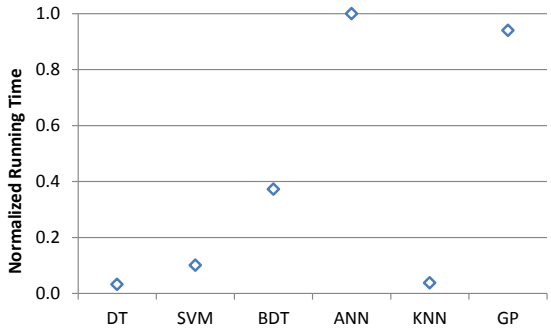
Testing Google Prediction API for Digit Recognition Problem

Performance (accuracy and running time) of Google Prediction API (version 1.6) are compared with several other supervised learning algorithms as shown in table and figures below. Accuracy of test data is defined as the ratio of correct prediction number over total test data samples. The machine learning algorithms include: Decision Tree (DT), SVM (Supported Vector Machine), ANN (Artificial Neural Network), Boosted Decision Tree (BDT) and KNN (K Nearest Neighbor). For each classifier, a 10-fold cross validation is implemented for model parameter selection in order to optimize performance. Running time measured recorded is the total time (seconds) spent in both training and testing.

Classifier	Model parameters	Accuracy on Test Data Set	Running Time (second)
Decision Tree (DT)	minimum sample split is 12 maximum depth is 10	0.66	3.3
SVM	kernel is 2-order polynomial	0.94	10.2
Boosted Decision Tree (BDT)	maximum depth is 1 number of estimator is 500	0.72	37.5
ANN	network size is 784-25-10	0.91	100.6
K-NN	k = 4	0.84	3.82
Google Prediction (GP)	-	**0.84**	**94.6**





The comparison indicates that accuracy of Google Prediction API is placed in the middle of other five algorithms. SVM and ANN have the highest accuracy. As for running time, google prediction API is more time consuming than other algorithms except Artificial Neural Network. No information is disclosed by Google as to which machine learning algorithm is used in Google Prediction API. As a generic prediction platform which isn't able to adjust its algorithms according to different problems, the performance of the API is ok.