Building a Student Intervention System: An Udacity Nanodegree ML Project

Omoju Miller

May 9, 2016

Introduction

Models

Decision Tree Classifier

- What is the theoretical $O(n)$ time & space complexity in terms of input size?
- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?

Table 1: Result of training with a DecisionTreeClassifier

	Training set size		
	100	200	300
Training time (secs)	0.001	0.001	0.002
Prediction time (secs)	0.000	0.000	0.000
F1 score for training set	1.000	1.000	1.000
F1 score for test set	0.683	0.703	0.758

Support Vector Machine

- What is the theoretical $O(n)$ time & space complexity in terms of input size?
- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?

Random Forest Classifier

- What is the theoretical $O(n)$ time & space complexity in terms of input size?

Table 2: Result of training with a Support Vector Machine

	Training set size		
	100	200	300
Training time (secs)	0.008	0.010	0.051
Prediction time (secs)	0.000	0.000	0.000
F1 score for training set	0.909	0.853	0.830
F1 score for test set	0.767	0.769	0.779

- What are the general applications of this model? What are its strengths and weaknesses?
- Given what you know about the data so far, why did you choose this model to apply?

Table 3: Result of training with a Random Forest Classifier

	Training set size		
	100	200	300
Training time (secs)	0.029	0.020	0.026
Prediction time (secs)	0.000	0.000	0.000
F1 score for training set	0.984	0.989	0.997
F1 score for test set	0.722	0.774	0.694

Conclusion

This paper has laid out some of the challenge of