Empirical Study of the Performance of Six Algorithms on Five Binary Classification Problems

Maximilian Siemers siemersm@gmail.com
Department of Cognitive Science
University of California, San Diego
San Diego, CA 92117, USA

Class: COGS 118A: Introduction to Machine Learning I
Instructor: Prof. Zhuowen Tu

Abstract

Keywords:

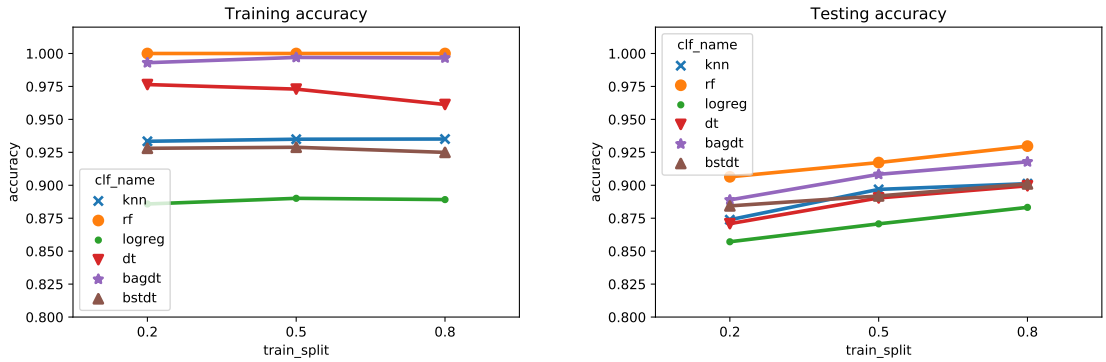
1. Introduction

2. Method

3. Experiment

3.1 Hyperparameters and Performance

3.2 Train/Test split and Performance



(a) Training accuracy (b) Testing accuracy

Figure 1: Training and testing accuracy by train split, averaged over problems and shuffles

Table 1: Classifier testing accuracy by problem, averaged over shuffles (0.8 train split)

Classifier	WDBC	INCOME	IRIS	COVTYPE	LETTER
bagdt	.968	.833	.944	.971	.873
bstdt	.968	.840	.956	.971	.771
dt	.953	.799	.956	.970	.820
knn	.950	.757	.933	.967	**.899**
logreg	.968	.792	**.967**	.968	.721
rf	**.982**	**.848**	.944	**.976**	.897

Table 2: Ranked classifiers by testing accuracy, averaged over problems and shuffles (0.8 train split)

Classifier	Accuracy
rf	.930
bagdt	.918
knn	.901
bstdt	.901
dt	.900
logreg	.883

3.3 Overall Classifier Performance

4. Conclusion

5. Bonus Points

6. References

Appendix A.