CSCI447: Analysis of Traditional Machine-Learning Algorithms on Real-World Data.

Christopher R. Barbour, John Sherrill, Brandon Fenton

September 3, 2015

Abstract

Hello, this is our abstractjfslfksjlfksdjflksdj lskjdfsljf;asdkfja;lkdjfl slkdfj;aksdjflksdjf a;sldkjfalskdfja;kljdf a;lksdjf;aklsjdfl ksjd.

Contents

1 Introduction

The purpose of this study is to assess the classification performance of 10 machine-learning algorithms on 10 datasets from the UCI repository of machine-learning datasets. The algorithms, listed in section 2.1, cover a broad spectrum of classification techniques with varying degrees of assumptions, computational intensity, and inductive biases. The latter of these will be the focus on our performance hypothesis of each dataset and interpretation of results.

2 Experimental Design

For each of the selected datasets and algorithms, we will randomly partition the examples into a training dataset which will be used to construct the classifier and a test dataset which will be used to test the classifiers ability to generalize, or correctly classify similar examples from the same population of interest. Each algorithm will be heuristically tuned to each dataset prior to the experiment using a single run 10-fold Cross-validation. The measures used to quantify this classification ability will be one minus the misclassification rate, the weighted average F-measure, and the weighted average Area under the ROC curve. These measures were chosen to assess different capabilities of each algorithm. Misclassification rate measures.... . Weighted Average F-measure related to .... . Weighted Average Area under the ROC assess.... .

2.1 Datasets

The datasets selected were chosen to contain a classification problem, be free of missing attributes or classes, and be of somewhat interest to the authors. Prior to the experiment, variables with each example having a unique value (i.e. a sample identification, or name of an animal) was removed. Table 1 displays the name of our datasets with brief descriptions, the types of attributes present, the number of classes for prediction, and the number of total examples.

2.2 Algorithm Descriptions

The algorithms implemented in the experiment are displayed below in Table 2. As stated earlier, individual tuning to each dataset was done prior to the experiment, and these results are displayed in the supplementary information. Some options remained fixed during all of the experiment, and these are discussed below along with certain examples of how tuning was performed.

For Naive Bayes estimation.. . For K nearest-neighbors..... For AdaboostM1.... For Ripper K.... For decision trees.... For the Feed-Forward Neural Network... For the Kernel Neural Network....

2.3 Hypothesis

3 Results

4 Discussion

5 References

<div align="center">

Supplementary Information

Tuning Parameter Information

Supplemental Plots of Results

</div>