Supervised Learning Algorithms: Classification Comparison

Howard Kim hokim@ucsd.edu
Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-5004, USA

Editor: Howard Kim

Abstract

This paper describes the replication of a large-scale empirical comparison between supervised machine learning algorithms of **?** in a smaller scale. For this particular comparison, four different supervised machine learning algorithms are compared and contrasted by their performance on binary classification tasks across four different data sets with five trials each. This process boils down to total of eighty different trials across the following algorithms: logistic regression, k-nearest neighbors, decision trees, and random forests. Each trial will be tuned by five-fold cross-validation on randomly chosen five-thousand data samples to select hyperparameters through a systematic grid searches. The results seem to be in parallel to the findings observed by **?** with performance measured through ACC, F1 score, and AUC.

Keywords: binary classification, logistic regression, k-nearest neighbors, decision trees, random forests, hyperparameters, cross-validation

1. Introduction

The precedent of this paper is the widely-known empirical comparison of supervised learning algorithms done by **?**, hereafter referred to as the CNM06. Much like this paper, the CNM06 also had a precedent

Much like the precedent to comparison and analysis done by **?**, hereafter referred to as CNM06,

Acknowledgments

We would like to acknowledge support for this project from the National Science Foundation (NSF grant IIS-9988642) and the Multidisciplinary Research Program of the Department of Defense (MURI N00014-00-1-0637).

Appendix A.

In this appendix we prove the following theorem from Section 6.2:
Theorem *Let u, v, w be discrete variables such that v, w do not co-occur with u (i.e., $u \neq 0 \Rightarrow v = w = 0$ in a given dataset $\mathcal{D}$). Let N_{v0}, N_{w0} be the number of data points for which $v = 0, w = 0$ respectively, and let I_{uv}, I_{uw} be the respective empirical mutual information values based on the sample $\mathcal{D}$. Then*

$$N_{v0} \; > \; N_{w0} \;\; \Rightarrow \;\; I_{uv} \; \leq \; I_{uw}$$

with equality only if u is identically 0. ■
Proof. We use the notation:

$$P_v(i) \; = \; \frac{N_v^i}{N}, \quad i \neq 0; \quad P_{v0} \; \equiv \; P_v(0) \; = \; 1 - \sum_{i \neq 0} P_v(i).$$

These values represent the (empirical) probabilities of v taking value $i \neq 0$ and 0 respectively. Entropies will be denoted by H. We aim to show that $\frac{\partial I_{uv}}{\partial P_{v0}} < 0$....

Remainder omitted in this sample. See http://www.jmlr.org/papers/ for full paper.

References