Supervised Learning

Yong-Lin Zhu[1, *]
[1]

(Dated: September 23, 2018)

Decision trees with some form of pruning 1. Learning curve: Error vs Training size 2. Tuning: Error vs max depth 3. pruning: cut some leaves Neural networks 1. Learning curve: Error vs Training size 2. Learn rates? 3. layer? 4. nodes? Boosting 1. Learning curve: Error vs Training size 2. n estimators 3. max depth More aggressive than DT 4. learning rate Support Vector Machines 1. Learning curve: Error vs Training size 2. At least two kernels

Keywords: Machine Learning, Supervised Learning

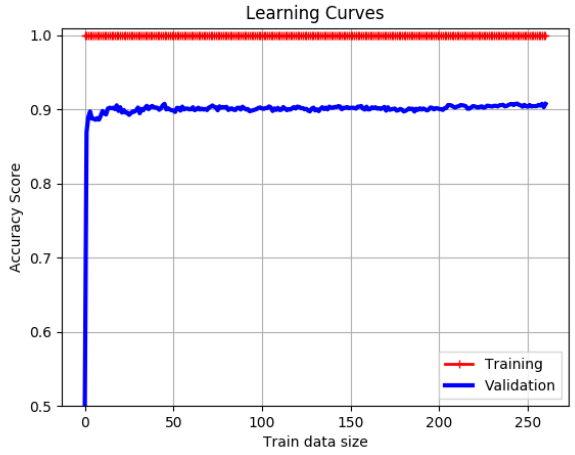

FIG. 1

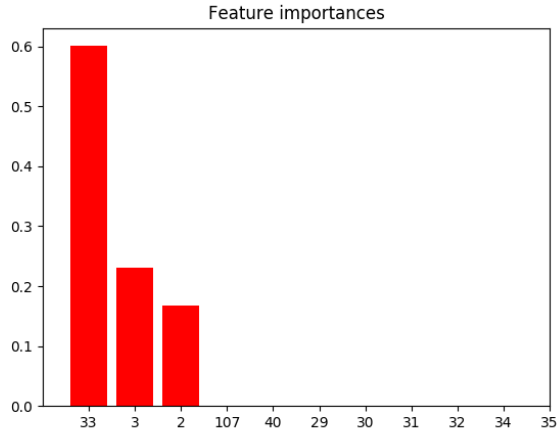

FIG. 2

Introduction Machine Learning is important and popular now.

There are many Supervised Learning algorithms. Discussions(or debates) on which algorithm is better never settles.(see interesting paper in [1]).

In this manuscript, I will five Supervised Learning algorithms on two dataset. The study includes the how to split train and test dataset, how to tuning the parameters to get better performance. The best-tuned performance of algorithms on two dataset are discussed at the conclusion.

Datasets Clean One hot encoding

DecisionTree

Neural networks

Boosting

Support Vector Machines

k-nearest neighbors

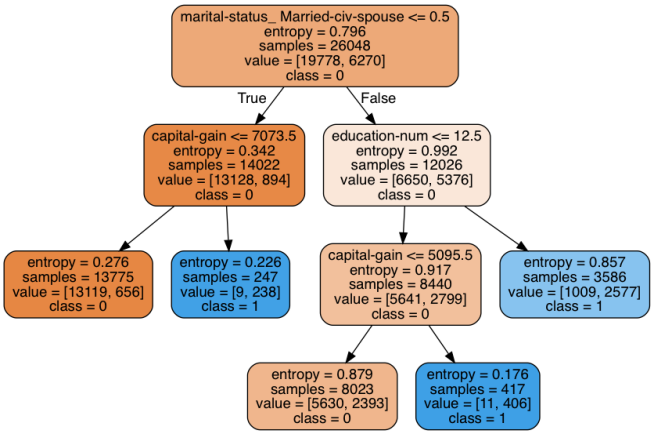

FIG. 3

* yzhu459@gatech.edu

[1] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, The Journal of Machine Learning Research

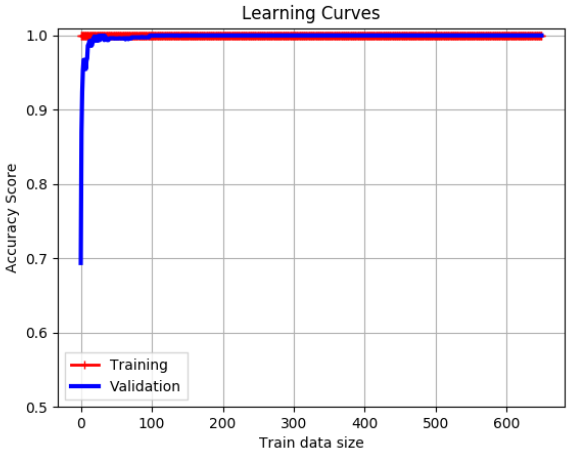

FIG. 4

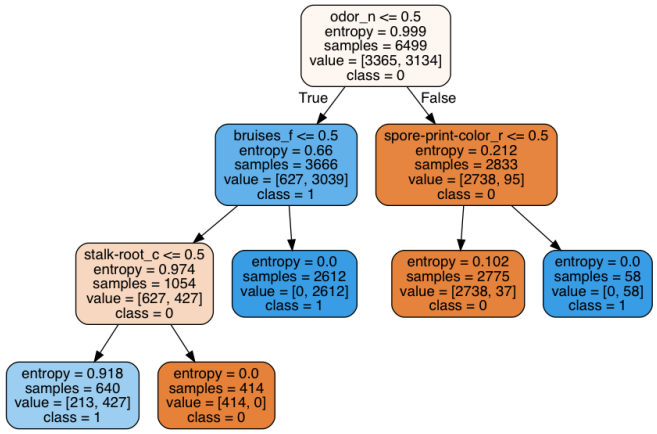

FIG. 6

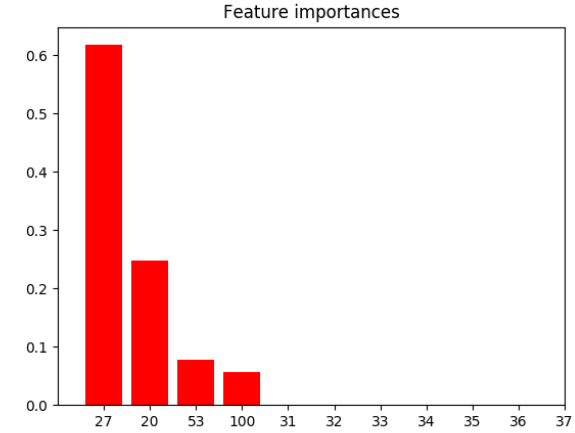

FIG. 5

15, 3133 (2014).