Supervised Learning Report

Datasets:

1. Banknotes

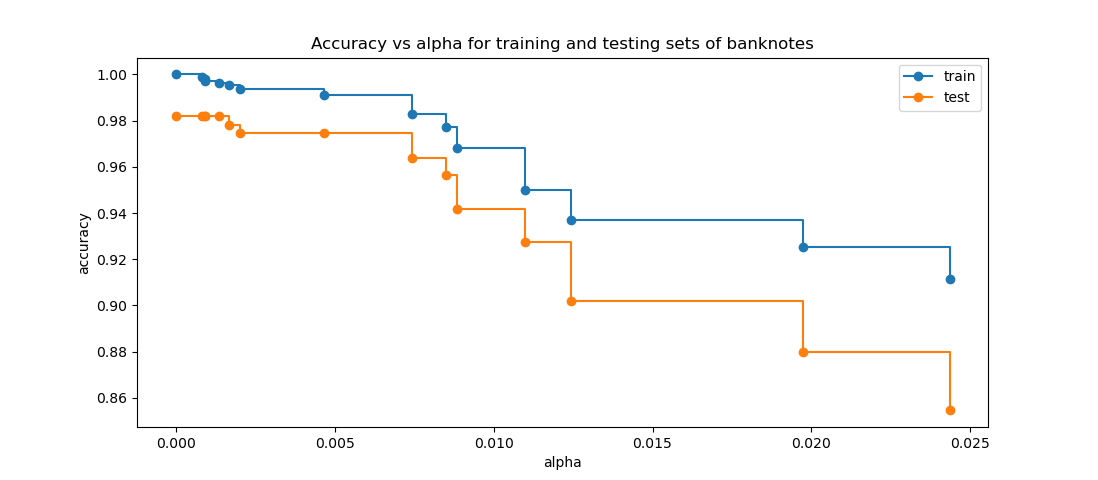
This is a binary classification dataset. Each instance consists of various measured attributes from a photo of a banknote. These attributes are different transforms of the image of each banknote. The classification of each instance is authentic or inauthentic, represented with the values 0 and 1, respectively.

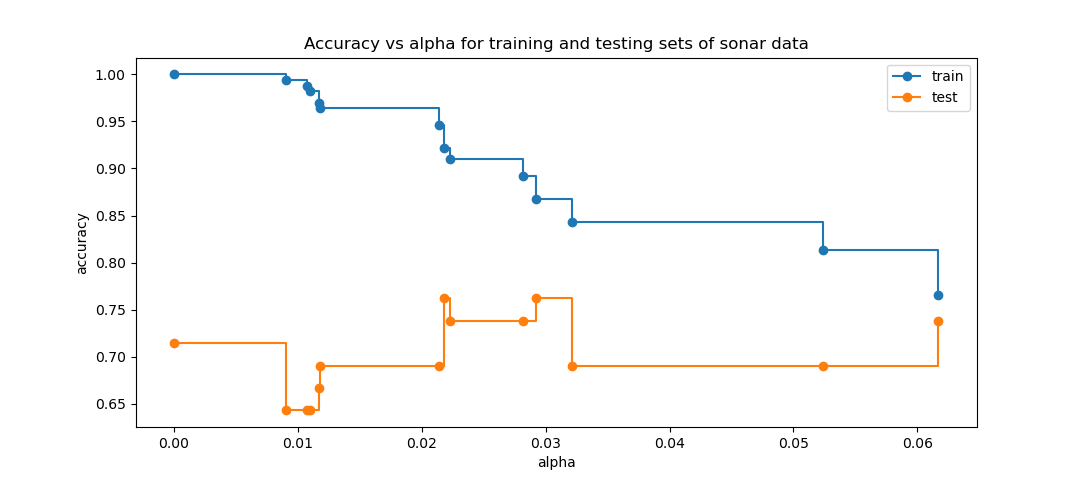
2. Sonar Dataset

This is also a binary classification dataset. The dataset addresses the problem of differentiating between rocks and mines using sonar data. Each instance consists of 60 sonar measurements from angles around the targets. The ‘mines’ used for testing were actually just metal cylinders. Each measurement consists of a value between 0 and 1 showing how strongly the sonar signal returned from the measured angles. The classification values are binary, ‘R’ for a rock and ‘M’ for a mine.

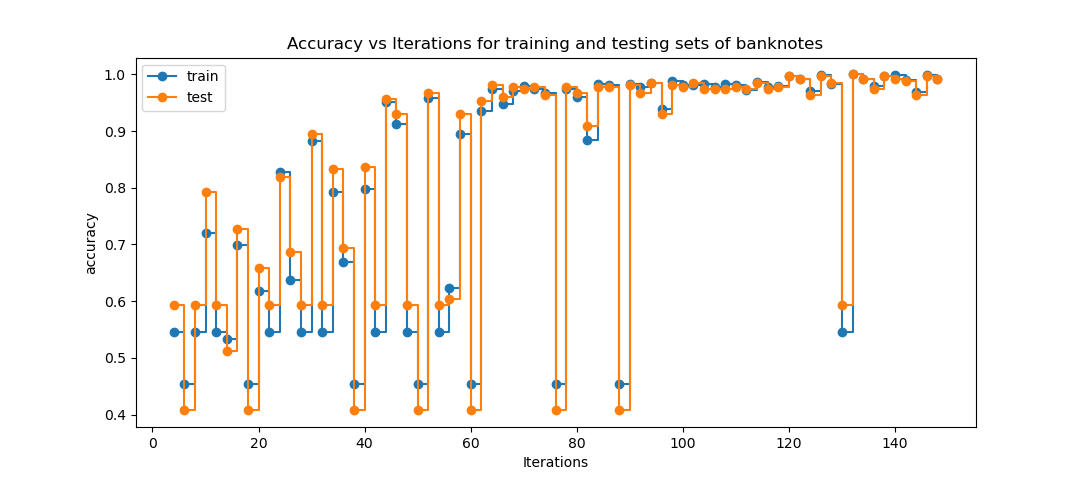
Both datasets are unbalanced with respect to their classifications. The sonar dataset has 111 observations of mines and 97 observations of rocks. The banknote dataset has 762 authentic banknotes and 610 inauthentic banknotes.

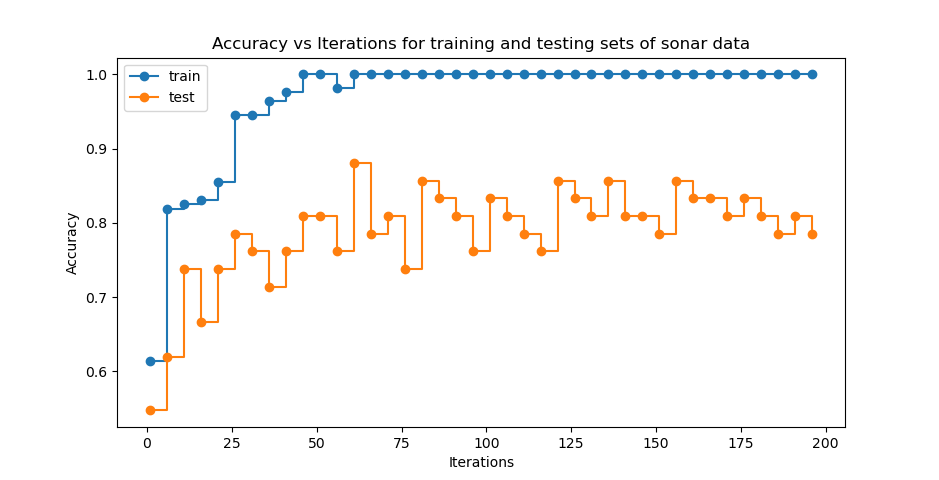
Decision Trees:



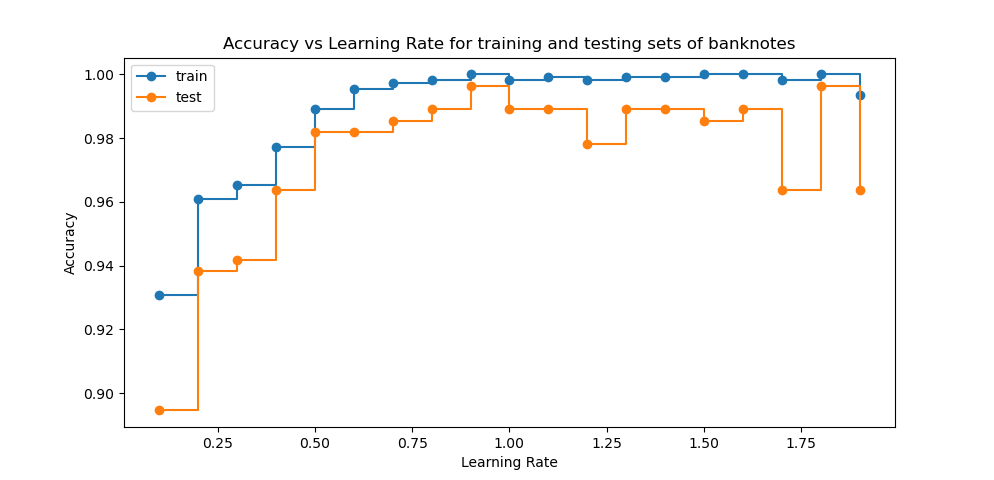


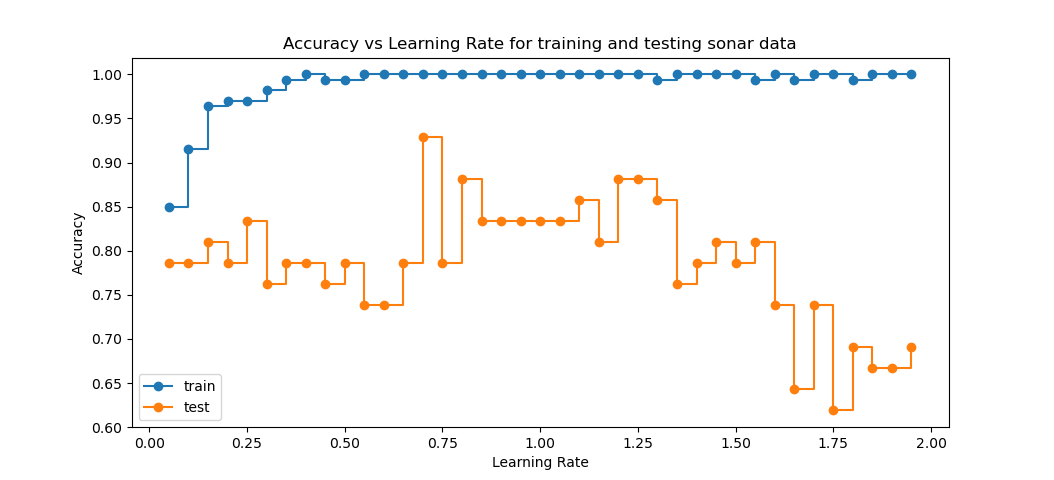
Neural Networks:



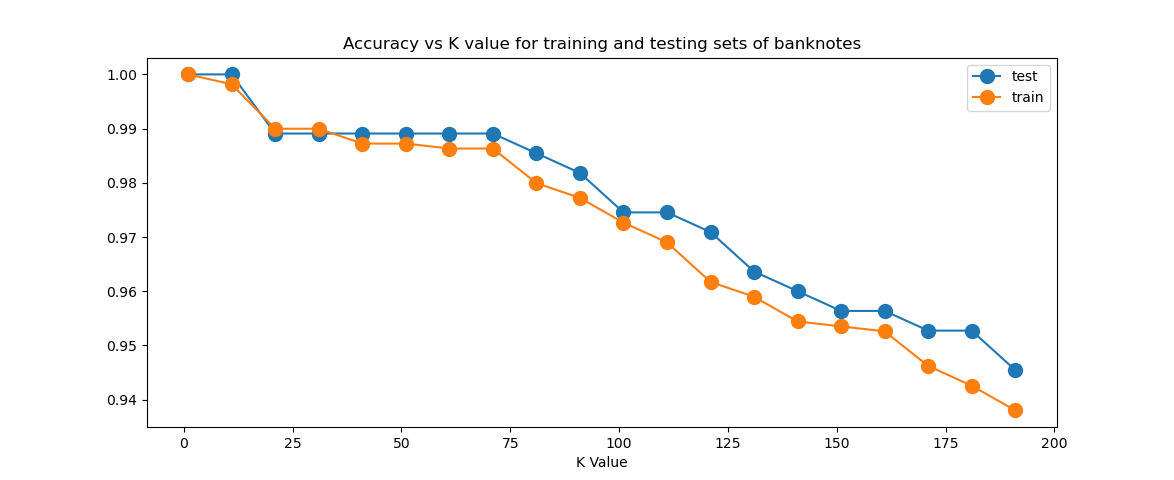


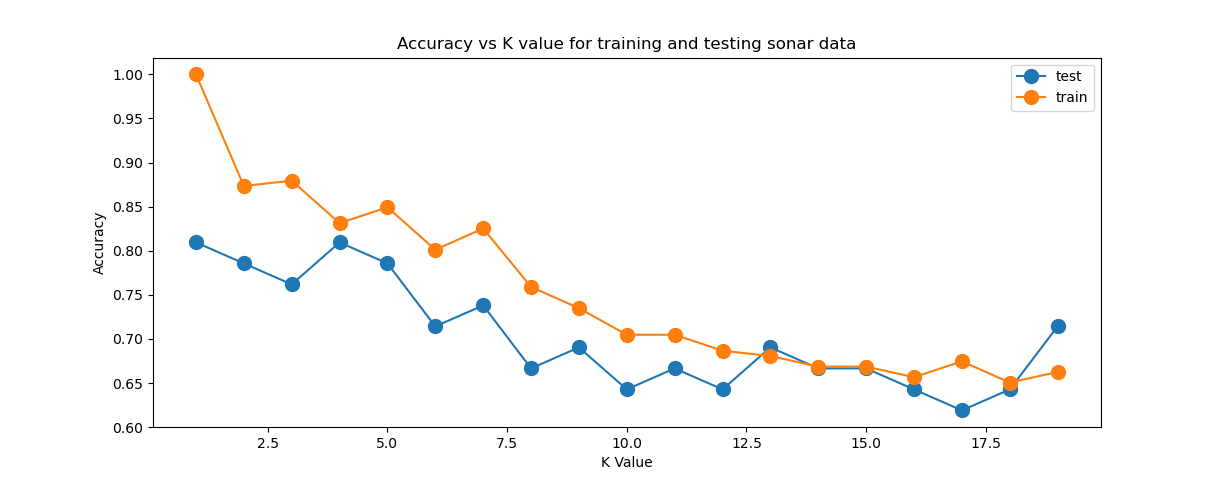
Boosting with decision trees:





K-Nearest Neighbors:





Support Vector Machines:

The banknotes were almost linearly separable, so a basic linear kernel for the applied SVM produced near perfect results, shown below:

Accuracy: 0.9927272727272727

Precision: 1.0

Recall: 0.984

Confusion Matrix:

[[150 0]

[ 2 123]]

The sonar data

Kernel | Exponent or Gamma Value | Training Score | Testing Score

poly 1 0.843373 0.666667

poly 2 0.891566 0.809524

poly 3 0.945783 0.857143

rbf 0.01 0.554217 0.595238

rbf 0.5 0.951807 0.833333

rbf 1.0 0.981928 0.904762

rbf 1.5 1.000000 0.904762