Simple binary classification

The Sonar Dataset involves the prediction of whether or not an object is a mine or a rock given the strength of sonar returns at different angles. It is a binary (2-class) classification problem. The number of observations for each class is not balanced. There are 208 observations with 60 input variables and 1 output variable. The variable names are as follows:

Sonar returns at different angles

…

Class (M for mine and R for rock)

The baseline performance of predicting the most prevalent class is a classification accuracy of approximately 53%. Top results achieve a classification accuracy of approximately 88%.

1. Split the data into training and test (consider class imbalance).
2. Choose a performance function suitable for the problem.
3. Adjust the following algorithms to the data using the default parameters: Linear Regression, Logistic Regression, SVM, KNN, Decision trees.
4. Select the algorithm with the highest performance measure.
5. Report the performance on the test data.

Database

https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.csv

More information

https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+(Sonar,+Mines+vs.+Rocks)