Multi-class Classification Using Different Algorithms

Jiacheng Liu * Zejing Wang † Naichen Shi ‡

April 2021

1 Data Set Description

- Link: http://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset
- This data set contains 13611 observations from 7 different types of registered dry beans, obtained by a computer vision system. There are 16 features to describe the data: 12 dimensions and 4 shape features. The goal of the research is to build a classification model that can predict the type of beans on the basis of the 16 features.
- The dataset has moderate size and multiple classes, thus extremely suitable for the study of multi-class classification algorithms.

2 Problem Description

The objective of this project is to compare the performances of different machine learning algorithms on the multiclass classification problem using the dry-bean dataset. The data set will be divided into the training set and test set. Each observation consists of a label (can be represented as the class number) and several features. Utilizing the features provided by the data set, we will perform a multi-class classification task on the training data set using different algorithms learned in class. For each algorithm, the model will be chosen by applying cross-validation on and the error rate results will be plotted graphically. The best-performed models from all algorithms will be selected by applying the training model to the test data sets and comparing the performances from different aspects. We will try to analyze the potential cause of performance differences for different algorithms and give some possible explanations on why the final selected algorithms are able to outperform others.

3 Statistical Learning Methods to be Studied

- Model-based Classification: Linear\Logistic Regression. These are basic classification algorithms we learnt in class.
- Tree-based Methods: CART, Bagging and Random Forests.
- Support Vector Machines: Support vector machine is popular and has good generalization performance.
- Neural Networks: Neural network is a parametric model that excels in many image classification task.

*First year graduate student from IOE, liujc@umich.edu
†First year graduate student from IOE, zejingw@umich.edu
‡First year graduate student from IOE, naichens@umich.edu