

## CSE 6363 - *Machine Learning*

Homework 2- Spring 2023

Due Date: April 23 2023

### Data Set Generation

This assignment consists of a number of implementation and result analysis questions. The problems in this assignment will use the data from Homework 1 Question 2 c).

Make sure that you enter your own student ID. Results on data for other student ID numbers will not be considered correct solutions.

### Ensemble Classifiers

1. Consider one more time the problem from Question 2 in the first assignment where we want to predict the type of material (among 3 material types) of a mug based on four measurements, namely the height, diameter, weight, and hue (color). Assume the same datasets you generated for the first assignment.
  - a) Implement a bagging routine for a softmax regression classifier. For the softmax classifier you can use your implementation from project 1 or utilize an existing implementation. If you use an existing implementation you need to reference where you obtained it in your writeup and instructions. You do need to implement the bagging routine yourself.
  - b) Apply bagging 10, 50, and 100 times to the training data. For each of the three cases, evaluate the resulting ensemble classifier on the test data set and compare the error rates for a single classifier and the three ensemble classifiers. Briefly discuss the results you obtained.
2. Using the data and softmax regression classifier, apply boosting to the problem.
  - a) Implement AdaBoost on top of your softmax regression classifier. Again, for the softmax classifier you can use and modify your version from Project 1 or modify an existing implementation to be able to work with weighted sample sets. If you use an existing implementation you need to reference where you obtained it in your writeup and instructions. You do need to implement the boosting routine yourself.
  - b) Apply boosting 10, 25, and 50 times to the training data. For each of the three cases, evaluate the resulting ensemble classifier on the test data set and compare the error rates for a single classifier and the three ensemble classifiers. Briefly discuss the results you obtained.

## **K-Means Clustering**

1. Consider an unlabeled version of the dataset for Problem 2 c) in the first assignment (i.e. assume that the labels are not part of the dataset).
  - a) Implement K-Means Clustering for this clustering problem.
  - b) Apply your K-Means clustering algorithm to the dataset with  $K=3$ , 6, and 9 and see how well the clusters align with the original label of the dataset. For this, compute the accuracy for each cluster assuming that the predicted label for each cluster is the most frequently occurring label in this cluster. Once you computed this for each cluster, compute the overall accuracy as the weighted sum of the accuracies where the weight is the fraction of data points in the corresponding cluster.