CSE 5243 – Introduction to Data Mining

Homework 2

Rajarshi Biswas, Sayam Ganguly

{biswas.91, ganguly.32}@osu.edu

The Ohio State University

Table of Contents

[Approach to train and test the Model 2](#_Toc474605504)

[Iris Dataset 2](#_Toc474605505)

[Income Dataset 2](#_Toc474605506)

[Performance of the classifiers 2](#_Toc474605507)

[Iris Dataset 2](#_Toc474605508)

[Income Dataset 2](#_Toc474605509)

[Comparison with off-the-shelf kNN implementation 2](#_Toc474605510)

# 1 Approach to train and test kNN Model

## Iris Dataset

We did not have to do any preprocessing on the Iris (train and test) datasets, other than normalizing the attribute values. All the values were numerical. Furthermore, there were no missing values in either the training or the test dataset.

## Income Dataset

Unlike the Iris dataset, Income dataset has missing values in both the training and testing dataset. Furthermore, the attributes of these datasets are .Hence, we have done the following preprocessing on the Income dataset.

### 1.2.1 Missing Values

We have 520 records in our datasets and among them there are 62 total missing values which are identified by the ‘?’ character.

### 1.2.2 Attribute Transformation

# Performance of the classifiers

## Iris Dataset

## Income Dataset

# Comparison with off-the-shelf kNN implementation