

CS 7267

MACHINE LEARNING

PROJECT 1

UNSUPERVISED LEARNING

#### INSTRUCTOR

**Dr. Zongxing Xie**

**Michael Rizig**

**001008703**

**1. ABSTRACT**

In this project, we are tasked with applying k-mean clustering to a dataset. This clustering method utilizes K number of averages, and groups the data into clusters based on their distance to the closest mean. By repeating k-mean clustering a few times, the clusters become more accurte and representative of their data. The data is then plotted to view the clustering accuracy and view the clear differences between clusters. Our kmtest dataset utilized 2 dimentional data, while the iris dataset utilizes 4 dimentional data. This means our program must adapt to any type of input data.

To view revison history and step by step building of this project view on my github:

**2. Test RESULTS**

**2.1 Clustering with K-means algorithm for kmtest dataset**

**...............................................................................................................................................**

**...............................................................................................................................................**

**...............................................................................................................................................**

**...............................................................................................................................................**

**...............................................................................................................................................**

|  |  |
| --- | --- |
| **(a)** | **(b)** |

**Figure 1:** (a) Original clustering, (b) K-means clustering result.

**2.2 Test Results for Clustering with K-means algorithm for iris dataset**

**This dataset is a 4 dimentional dataset.**

**Figure 2.2.a:** This 3 dimentional graph shows 3 of the 4 dimentions of each datapoint. Each group or ‘cluster’ is identifierd by their unique color.

**Figure 2.2.b: This**

|  |  |
| --- | --- |
| **(a)** | **(b)** |

**Figure 2:** (a) Original clustering, (b) K-means clustering result.

**3. CODES**

**3.1 Code for K-means algorithm for kmtest dataset**

# Name: John

# Number: 123456

# Project 1

# Dataset: iris

**3.2 Code for K-means algorithm for iris dataset**

# Name: John

# Number: 123456

# Project 1

# Dataset: iris