Probability Course Grading Report

Karthikeya Hanu Prakash Kanithi

November 20, 2023

1 Introduction

This report outlines the grading process for the course 'EE23010: Probability and Random Processes' in Fall 2023, offered by Dr. G V V Sharma using k-means clustering. The goal is to assign grades based on student performance, as determined by clustering their marks.

2 Data and Preprocessing

The marks data was obtained from the course assessments and stored in an Excel file. The Python script used for clustering reads the data, standardizes it, and applies k-means clustering with seven clusters.

3 K-Means Clustering

3.1 Overview

K-means clustering is an unsupervised machine learning algorithm used for partitioning a dataset into K distinct, non-overlapping subsets (clusters). The algorithm works iteratively to assign each data point to one of K clusters based on its features.

3.2 Algorithm Steps

1. **Initialization:** Choose K initial cluster centroids randomly.
2. **Assignment:** Assign each data point to the nearest centroid, forming K clusters.
3. **Update:** Recalculate the centroids based on the mean of data points in each cluster.
4. **Repeat:** Repeat steps 2-3 until convergence or a maximum number of iterations.

4 Grading Based on Clusters

The resulting clusters are mapped to grades as follows: (CLusters are randomly generated)

- **Cluster 0:** B-
- **Cluster 1:** A-
- **Cluster 2:** D
- **Cluster 3:** B
- **Cluster 4:** A
- **Cluster 5:** C-
- **Cluster 6:** C

Each student is assigned a grade based on the cluster to which their performance belongs.

5 Conclusion

This grading process provides a systematic way to categorize student performance in the Probability course. The use of k-means clustering helps in objectively determining grade boundaries.