EEE311 Introduction to Machine Learning

K-means Clustering

Step 1: Data Generation

- Generate three sets of 2D data from Gaussian distributions with means at [0,0], [3.5,3.5], and [-3.5,3.5] and unit covariance.
- Ensure each group has 400 samples.

Step 2: Initialization

- Randomly select three points from the dataset as initial cluster means.

Step 3: Cluster Assignment

- Calculate the Euclidean distance of each point to the cluster means.
- Assign each point to the nearest cluster.

Step 4: Update Cluster Means

- Recalculate the means of each cluster based on the assigned points.

Step 5: Iterative Optimization

- Repeat Steps 3 and 4.
- Continue this iterative process until fewer than 20 samples change their cluster assignment in a single iteration.

Step 6: Visualization through Video/Demo

- Write a script in Python that captures the state of the clusters at each iteration.
- Use a plotting library (e.g., Matplotlib) to visualize the clusters.