Aditya Patel

STAT 598 – Statistical Methods in Machine Learning

Homework 3

1. PCA (Left), LDA (Right)

A graph of a graph of a graph

Description automatically generated with medium confidence

b)

Define points as a matrix:

Calculate Means and center the data:

Define Covariance Matrix:

Eigenvalues and Eigenvectors:

Projecting the points into the 1-d space:

**The variance of the projected data is equal to the eigenvalue of the principal component. Therefore, the variance of the projected data is 2.**

Reconstructing the data into 2-d spaces

Reconstruction error:

1. K-Means Clustering
2. Cluster Centers:

A group of black squares with white letters

Description automatically generated

1. Accuracy Metrics:

A black screen with white text

Description automatically generated

1. Inertia vs. Cluster Numbers:

A graph with a line

Description automatically generated

By the elbow method, we pick K when at the point where the inertia stops dropping rapidly. As evidenced in the plot of the k-value vs. inertia, we see that the best k-value to pick is indeed 4.

The inertia drops from 1.6e11 to 1.3e11 from k=1 to k=4, and then from k=1.3e11 to 1.1e11 between k=4 and k=10, which means that the best k-value is 4.

1. GMM Monte Carlo
2. A graph with orange dots

   Description automatically generated
3. The MMD score was identified as 0.0021.