CS215 Assignment-2

210050115 Patil Vipul Sudhir

210050119 Hari Prakash Reddy

1 PCA and Hyperplane Fitting

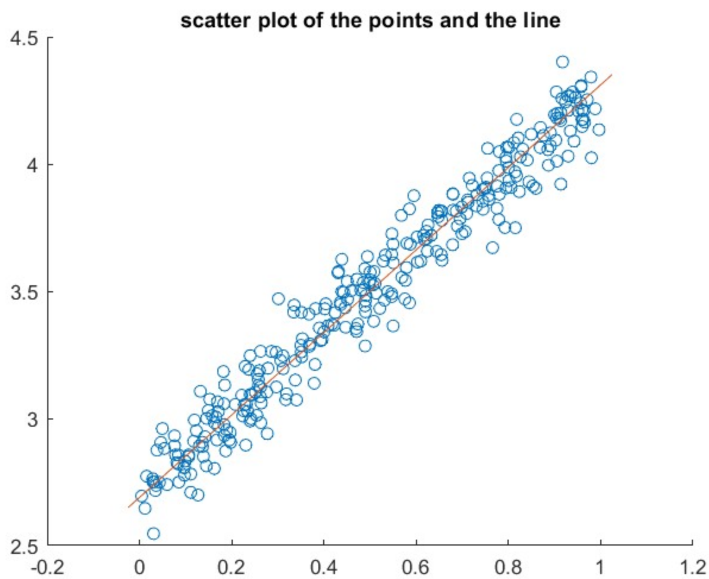
1.1 Description

The objective of PCA is to minimize the number of dimensions with significant weight.
Procedure:

1. Took the whole dataset consisting of 2 dimensions and made changes to our new dataset become 1 dimensional.
2. Computed the mean for the whole dataset so that it gives a point
3. Computed the covariance matrix of the whole dataset
4. Computed eigenvectors and the corresponding eigenvalues
5. Chosen 1 eigenvector with the largest eigenvalue to form a 2 x 1-dimensional matrix W
   - PCA simply performs a rotation of the coordinate axis and we change that coordinate axis to W

1.2 Q2b

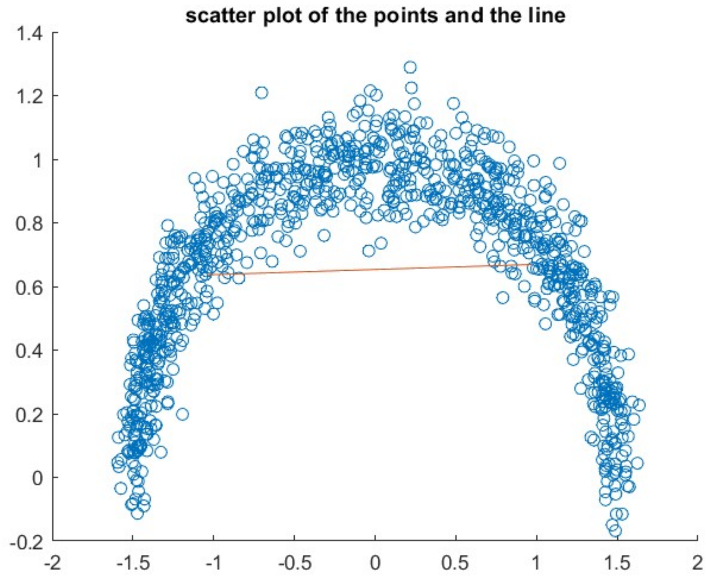
scatter plot of the points and the graph of a line showing the linear relationship between Y and X



Instructions to run the code:
Run q3b.m from code folder of Q3

1.3 Q2c

1.3.1 Plot



Instructions to run the code:
Run q3c.m from code folder of Q3

1.3.2 Comparison

For data set 1 the resulting approximation due to the PCA fits exactly into the scatter plot, but for 2nd data set it doesn't fit the way it fits for data set 1, the reason being that the data set1 is linear and data set 2 is non-linear.

1.3.3 Justification

PCA simply performs a rotation of the given coordinate axes. Rotation is a linear operator. But the axis frame returned by PCA (the eigenspace) captures every bit of the variance of the original data. Generally, we drop axes with small eigenvalues (weights) and proceed on a smaller data set. If data has a nonlinear structure, then PCA will have a larger number of dimensions with nonzero weights.

If the dataset is linear, then dropping any eigenvector doesn't cause much tampering as the variance mostly stays in direction of one of the eigenvectors, and dropping the lower one doesn't cause much tampering, but if the data set is non-linear then both the resulting eigenvalues has comparable eigenvalues(variance along eigenvectors), dropping one of them causes the resulting approximation to be more error-prone as it is ignoring a major variance along the dropped eigenvector.