IMLPR Homework #10

1. K-means Clustering

Given the following data x_1=(1,0), x_2=(1,1), x_3=(0,1), x_4=(0,0), x_5=(3,1), x_6=(4,1), x_7=(4,0), x_8=(3,0), x_1 through x_9, try k-means clustering by hands with k=2 and Euclidian distance. You will need to initialize the centroid.

On the same data, run the Matlab k-means and compare the results. Try the Matlab silhouette plot for clustering evaluation.

2. PCA

Given the following data points, (2,1), (2,4), (4,1), and (4,3)

(1) Project the data on the x-axis indicated by (1,0). Then compute the mean and variance of the projected data

(2) Project the data on the y-axis indicated by (0,1). compute the mean and variance of the projected data

(3) Now project the data on the axis indicated by $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. compute the mean and variance of the projected data

(4) Now compute the covariance matrix of the original data and find the eigenvalues and eigenvectors. Project the original data to the first principal component axis. Compute the mean and variance of the projected data.

(5) Based on the projection results of (1)~(4), which is the best projection result? Discuss the projection with respect to the variances? Plotting the original and projected data should help your understanding.