Michael Lee

HW #4

1. **PCA by hand**
2. First we write the matrix **X** of our data

X =

1. We calculate the mean of and subtract from each column

\*Both averages are 0 so subtracting them from the columns makes no difference

1. Compute the covariance matrix

The covariance matrix is

1. Find the eigenvalues and eigenvectors

We solve for the eigenvector corresponding to

Now we row reduce the matrix

We solve for the eigenvector corresponding to

Now we row reduce the matrix

Since corresponds to a greater eigenvector than it will be the first principal component .

The variance of

1. If we reconstruct the original points by using reconstruction error formula:

Answer: The reconstruction error is zero.

**2 Manual Calculations of one round of EM For a GMM**

**M-Step**

1. The objective function of each parameter is:

So the maximum likelihood is given by:

1. Computations for

1. Computations for
2. Computations for

**E-Step**

1. To calculate the new values of the expectation I used matlab to ensure my calculations were correct. After performing the E-step I got the following points for R:

So

**3 Code Write-Up**

1. The results matched my expectations, that with increased k we would see a smaller mistake rate and within-group sum of squares decrease. I thought that with more clusters we would observe less error since each observation would be closer to a center (since there would be more centers as the number of clusters increases).