**Machine Learning Programming Assignment 5**

**Comp540 Spring 2015 due 3 April 2015 at 8 pm**

MW56&HZ32

**Problem 1: k-means clustering (25 points)**

You will implement the two phases of the k-means algorithm separately in the next sections.

**Finding closest centroids (10 points)**

Your task is to complete the code in findClosestCentroids.m. You can implement this using a loop over every training example and every centroid.

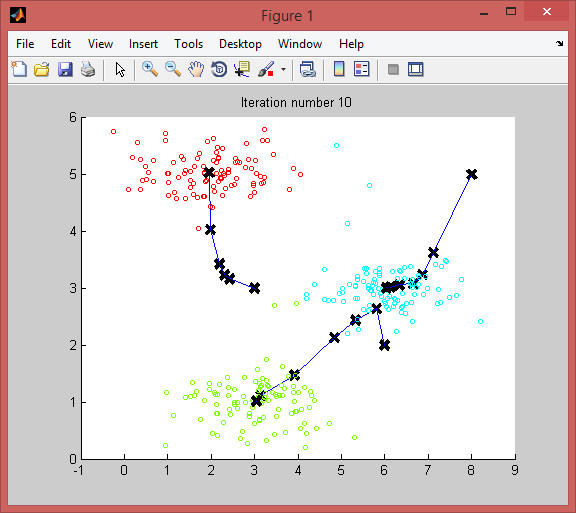
When you run the script ex5.m you should see the output [1 3 2] corresponding to the centroid3 assignments for the first 3 examples in our data set.

**Computing centroid means (10 points)**

You should now complete the code in computeCentroids.m. You can implement this function using a loop over the centroids.

**k-means on example dataset**

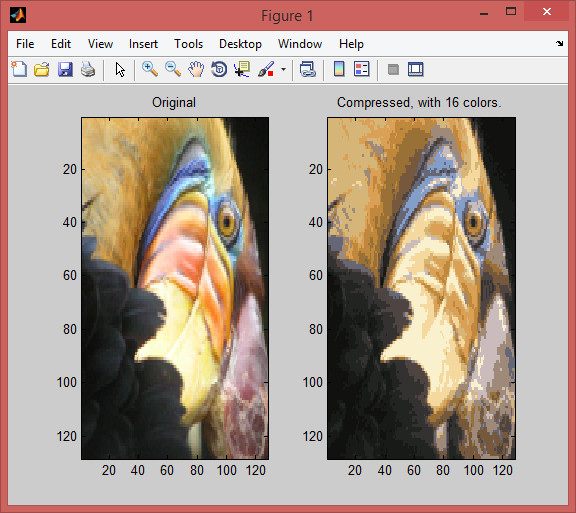
After you have completed the two functions (findClosestCentroids and computeCentroids), the next step in ex5.m At the end, your figure should look as the one displayed in Figure 1.



**Random initialization (5 points)**

, you should complete the function kMeansInitCentroids.m

**Image compression with k-means**

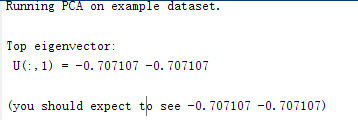


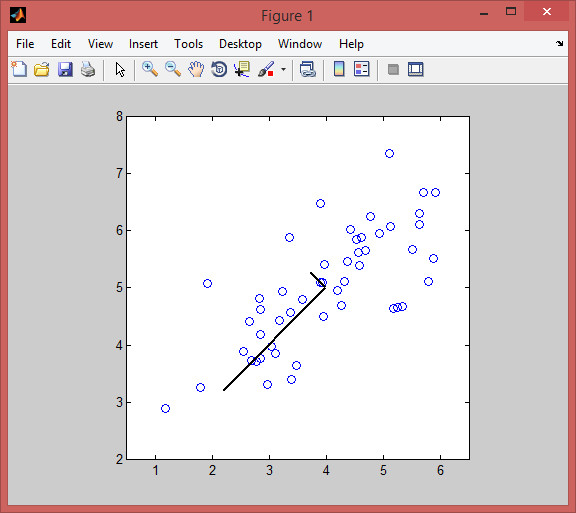
**Principal Components Analysis (15 points)**

**Implementing PCA (5 points)**

you should expect to see an output of about [-0.707

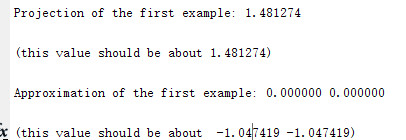
-0.707].





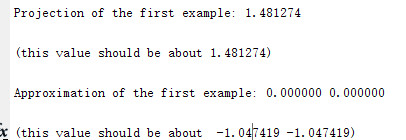
**Projecting the data onto the principal components (5 points)**

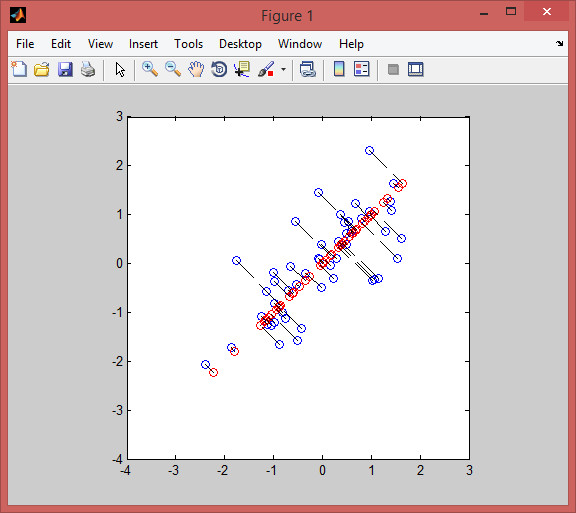
you should see a value of about 1.481 (or possibly -1.481, if you got *−U*1 instead of *U*1).



**Reconstructing an approximation of the data (5 points)**

you should see a value of about [-1.047 -1.047].



**Visualizing the projections**

**Face image dataset**

**PCA on faces**

If you want, you can also change the code to display more principal components to see how they capture more and more details.

**Dimensionality reduction**

