Week 3:

1a. The data needs to be standardized before running K-Mean clustering. Normalization standardizes the data.

1b. I would assign numeric values to categories or binarize (if applicable).

1c. Clustering algorithms cannot be influenced nor corrected as opposed to supervised learning algorithms that can be directed.

2. The call from within simpleKMeans to simpeDetermineCentroids(points, clusterID, size(centroids, 1)) returns closest centroid matrix.

3. The call from within simpleKMeans to

Given the following: simpleDetermineCentroids determines centroid for cluster 2 by finding the mean of all points that belong to cluster 2. How does simpleKMeans know which returned centroid is the one for cluster 2? (Answer in one sentence or less by describing the data structure)

Question 4:

Normalization in simpleKMeansFinished.m - Done

Question 5:

Write answers to item 1, 2, & 3 into the completed simpleKMeansFinished.m and submit to Catalyst by Saturday 11:00 PM.