In this assignment I was must to create two functions. First is for finding the closest centroid, second is for computing the mean.

```
In [1]: import scipy.io
        import os
        import numpy as np


        data = scipy.io.loadmat(os.path.join('ex7data2.mat'))
        X = data['X']
        initial_centroids = np.array([[3 ,  3] , [6 , 2] , [8 , 5]])
```

I import all needed libraries and download the data. Here I have initial centroids with known values.

1) Find closest centroids function:

After for loop, where I have “distances” variable, I compute the distance between training example and centroids.The distance is “eucledian”. By “argmin” I find the closest Centroid. Then by return I retirn the index of the closest centroids.

```
In [2]: from sklearn.metrics import pairwise_distances
        from sklearn.preprocessing import normalize

        def findClosestCentroids(X , centroids ):
            closest = np.ones((len(X), 1), dtype= np.uint8)
            for i in range(len(X)):
                distances = (centroids - X[i])**2
                distances = np.sum(distances, axis=1)
                argmin = np.argmin(distances)
                closest[i] = argmin

            return closest

        idx = findClosestCentroids(X , initial_centroids )
        print(idx[ :3])
```

```
[[0]
 [2]
 [1]]
```

2) Function for computing the mean:

First I create the new centroids matrix, which called “centroids”. In the for loop, I find the trainig sets with same class. At the end I compute the mean of the points. Then I return new moved centroids.

```
In [3]: def computeCentroids(X, idx, K):
            centroids = np.zeros( (K, X.shape[1]) )
            for k in range(K):
                Ck = [i for i in range(idx.size) if idx[i] == k]
                centroids[k,:] = np.mean( X[Ck,:], axis = 0 )

            return centroids
```

Then I call this function where K(number of clusters) is equal to the 3.

The result our centroids:

```
In [4]: K = 3
        idx = findClosestCentroids(X,initial_centroids)
        centroids = computeCentroids(X, idx, K)
        centroids
```

```
Out[4]: array([[2.42830111, 3.15792418],
               [5.81350331, 2.63365645],
               [7.11938687, 3.6166844 ]])
```

Then I run the following code which was given in the doc:

```
In [5]: import matplotlib.pyplot as plt
        max_iters = 10
```

```
for i in range(max_iters):
    idx=findClosestCentroids(X, centroids)
    centroids=computeCentroids(X, idx, K)
```

In order to plot the result I use matplotlyb library.

```
In [6]: centroids=initial_centroids
        plt.figure(figsize=(12,8))
        plt.scatter(X[:,0], X[:,1], c=idx, marker='8', s=20)
        plt.scatter(centroids[:,0], centroids[:,1], marker='+', s = 500, c='blue')
        plt.show()
```

The result is here: