[50 points]

In this coding assignment you are to implement the K-means unsupervised learning algorithm. Use the simplified Iris dataset to test your code.

- Column 1: sepal length
- Column 2: sepal width
- Column 3: class label (1= setosa, 2=versicolor)
- 50 samples for setosa, 50 samples for versicolor

Instructions:

1. Read “simple_iris_dataset.dat”
2. Create two clusters using K-means algorithm
3. Plot the clustered data with two different colors
4. Plot the centroid of each cluster on the same plot
5. Since ground truth is available, you can evaluate the accuracy of the algorithm. Show the confusion matrix.
6. Show the number of iterations required for the algorithm to converge
7. Run the algorithm several times

Hints:

- Take 2 samples randomly from the dataset to initialize the centroids
- The cluster assignment is random. When forming the confusion matrix, swap the off-diagonal elements with the diagonal elements if they are larger.
- To plot clustered data with two different colors:

```
figure;
hold on;
xlabel('Sepal Length');
ylabel('Sepal Width');
plot(X(idx_c1,1),X(idx_c1,2),'r.','MarkerSize',12)
plot(X(idx_c2,1),X(idx_c2,2),'b.','MarkerSize',10)
```

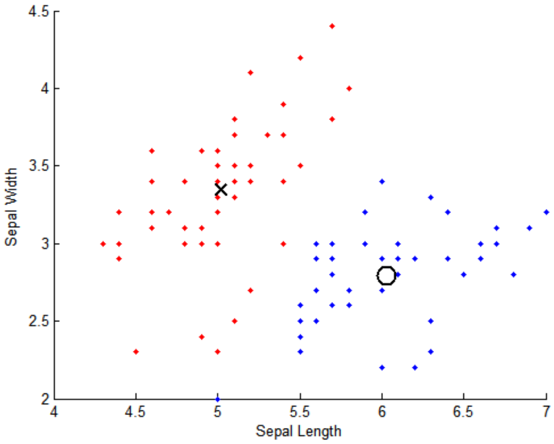
```
where idx_c1 is the idx_c1-th sample assigned to cluster 1; idx_c2 is the
idx_c2-th sample assigned to cluster 2
```

- To plot centroid of each cluster on the same plot:

```
plot(ctr1(:,1),ctr1(:,2), 'kx', 'MarkerSize',12,'LineWidth',2);
plot(ctr2(:,1),ctr2(:,2), 'ko', 'MarkerSize',12,'LineWidth',2);
```

```
where ctr1 and ctr2 are the cluster 1 centroid and cluster 2 centroid,
respectively.
```

Expected output:



```
confusion_matrix =

    50     0

     5    45

Convergence was achieved after 7 iterations
```