

Lesson 07 Quiz

Started: Feb 25 at 12:25pm

Quiz Instructions

Unsupervised Learning with K-Means

[Overview](#) | [Labs](#) | [Assignment](#) | [Code Talk](#) | [Quiz](#)

This quiz refers to **K-Means Clustering**.

You are allowed 2 attempts; your highest score will be kept. Correct answers will be shown after the 2nd attempt.

We need to modify the following K-means script so that it can **accept predefined centroids** (input variable C). Line 1 will need to change in order to ensure that the new version of the script accepts centroids as inputs.

```
01 def kmeans(X, k, th):
02     if k < 2:
03         print('k needs to be at least 2!')
04         return
05     if (th <= 0.0) or (th >= 1.0):
06         print('th values are beyond meaningful bounds')
07         return
08
09     N, m = X.shape # dimensions of the dataset
10     Y = np.zeros(N, dtype=int) # cluster labels
11     C = np.random.uniform(0, 1, [k,m]) # centroids
12     d = th + 1.0
13     dist_to_centroid = np.zeros(k) # centroid distances
14
15     while d > th:
16         C_ = deepcopy(C)
17
18         for i in xrange(N): # assign cluster labels to all data points
19             for j in xrange(k):
20                 dist_to_centroid[j] = np.sqrt(sum((X[i,] - C[j,])**2))
21             Y[i] = np.argmin(dist_to_centroid) # assign to most similar cluster
22
23         for j in xrange(k): # recalculate all the centroids
24             ind = FindAll(Y, j) # indexes of data points in cluster j
25             n = len(ind)
26             if n > 0: C[j] = sum(X[ind,]) / n
27
28         d = np.mean(abs(C - C_)) # how much have the centroids shifted on average?
29
30     return Y, C
```

Question 1**1 pts**

Apart from changing the inputs of the function, which modification makes sense to do inside the function to make the new script operational?

- ☒ Line 11 needs to be removed
- ☐ Line 10 needs to be removed
- ☐ Lines 10, 11, and 16 need to be removed
- ☐ Line 16 needs to be removed

Question 2**1 pts**

How does removing the line affect the function of K-means?

- ☐ K-means will be faster
- ☐ K-means will be slower
- ☒ K-means will be more chaotic
- ☐ K-means will be slower and more chaotic.
- ☐ K-means will be faster and more chaotic.
- ☐ K-means will be deterministic

Question 3**1 pts**

Will you need to do sensitivity analysis on the results, if you were to modify the script as instructed?

- ☐ Sometimes, but not so often
- ☐ No
- ☒ Yes

Question 4**1 pts**

Do not use k-means for this question, just a plot of the data.

points: (1,1), (1,2), (2,1), (2,2), (10,1), (11,2), (12,1),
(12,2), (1,11), (1,12), (2,11), (2,12)

How many clusters are there in this distribution of points, based on your observation of the data?

- ☐ 2 clusters
- ☐ More than 4 clusters
- ☐ 1 cluster
- ☒ 3 clusters

Question 5**1 pts**

Consider the following points:

(4, 10), (2, 4), (6, 4)

What's the centroid of the 2-D dataset's cluster comprising of the above points?

- ☒ (4, 6)
- ☐ (6, 4)
- ☐ (6, 6)
- ☐ (4, 4)

Question 6**1 pts**

Try clustering a homogeneous data set like the following (use any cluster centers you like):

points: (1,1), (1,2), (2,1), (2,2), (0,0), (0,1), (1,0), (1.5, 1.5),
(0.5, 0.5), (1.5, 0.5), (0.5, 1.5)

How does the choice of K affect the result?

- ☐ None of these
- ☒ The results change significantly if you were to use K = 3 instead of K = 2
- ☐ The results remain the same regardless of what value of K you use
- ☐ The results change significantly if you were to use K = 4 instead of K = 3
- ☐ The results change slightly if you were to use K = 3 instead of K = 2

Question 7

1 pts

Use the modified script from Question 1

```
centerGuesses: (1,1), (2,2), (11,1), (1,11)
points: (1,1), (1,2), (2,1), (2,2), (10,1), (11,2), (12,1),
        (12,2), (1,11), (1,12), (2,11), (2,12)
```

How many clusters will K-means find in this distribution of points given the cluster center guesses?

- ☒ 4 clusters
- ☐ 3 clusters
- ☐ 1 clusters
- ☐ 2 clusters

Question 8

1 pts

When would you predefine the centroids in K-means? Why?

- ☐ It's irrelevant
- ☐ Never, because your guess is as good as anyone else's
- ☐ When you want to avoid sensitivity analysis afterwards
- ☒ When you have a clear idea of what you expect from the algorithm

Not saved

Submit Quiz