

Feedback — Assignment 7

[Help](#)

You submitted this quiz on **Sat 7 Dec 2013 4:00 PM PST**. You got a score of **9.00** out of **10.00**.

In this assignment, you need to conduct clustering on data set asgn7-clustering.

This is not a real data set, but is simulated data, generated for the purpose of this assignment.

Please conduct this assignment in RapidMiner.

Question 1

Conduct k-means clustering with $K=2$, using k-Means. (Not W-SimpleKMeans). Use factory settings for all other parameters. Drag arrows between both of the output ports of k-Means to the results at the right. Look at the Centroid Table within the Cluster Model tab. Which attributes have the biggest difference between cluster_0 and cluster_1?

Your Answer	Score	Explanation
<input type="radio"/> C and F		
<input checked="" type="radio"/> A and F	✓ 1.00	
<input type="radio"/> C and E		
<input type="radio"/> B and D		
Total	1.00 / 1.00	

Question 2

Now look at the Plot view in the Example set tab. Set the X axis to be the first answer from Question 1, and set the Y axis to be the second answer from Question 1. (i.e. if you chose G and H, set X to G, and Y to H). Now set the color column to cluster. There are seven major groupings (“lumps”) in this data. How many of them are red?

You entered:

1

Your Answer		Score	Explanation
1	✓	1.00	
Total		1.00 / 1.00	

Question 3

What did k-Means do here?

Your Answer		Score	Explanation
<input checked="" type="radio"/> It found the least central lump in the data and made it a cluster	✓	1.00	
<input type="radio"/> It did a median split on two key variables			
<input type="radio"/> It split the lumps in the data approximately evenly in clusters			
<input type="radio"/> It found the most central lump in the data and made it a cluster			
Total		1.00 / 1.00	

Question 4

Now re-run k-Means with k=7. Did k-Means find the 7 lumps in the data that you saw earlier?

Your Answer		Score	Explanation
<input type="radio"/> Yes			
<input checked="" type="radio"/> No	✓	1.00	
Total		1.00 / 1.00	

Question 5

Plot each of the other variables against each other (not including the variables in your answer to question 1). Is there meaningful structure in any of these variables?

Your Answer	Score	Explanation
<input checked="" type="radio"/> No	✓ 1.00	
<input type="radio"/> Yes		
Total	1.00 / 1.00	

Question 6

Filter out all of the variables except the ones in question 1, and re-run k-Means using just these two variables. Is each of the seven data lumps now incorporated into its own reasonable cluster?

Your Answer	Score	Explanation
<input type="radio"/> Yes		
<input checked="" type="radio"/> No	✓ 1.00	
Total	1.00 / 1.00	

Question 7

What happened?

Your Answer	Score	Explanation
<input type="radio"/> Several clusters were devoted to regions of space without a lump		
<input type="radio"/> Two regions of space without lumps got clusters, and three lumps got a single cluster		
<input type="radio"/> It looks the same as when k= 2		
<input checked="" type="radio"/> One region of space without a lump got a cluster, and three	✓ 1.00	

lumps got split across two clusters

Total	1.00 /
	1.00

Question 8

Try running Expectation Maximization Clustering with $k=7$. Look at the cluster probabilities for each cluster in Plot View. Which clusters (if any) include lots of outliers?

Your Answer	Score	Explanation
<input checked="" type="radio"/> Cluster 2 and cluster 6	✓ 1.00	
<input type="radio"/> Cluster 5		
<input type="radio"/> There are no outlier clusters		
<input type="radio"/> Cluster 4		
Total	1.00 / 1.00	

Question 9

For fun, you might notice that outliers close to the center of the top-right cluster still got placed into an outlier cluster. This is the power of having centers and radii. When you're done looking, try running Agglomerative Clustering. Look at the Dendrogram. Nifty, huh?

Your Answer	Score	Explanation
<input checked="" type="radio"/> Yes, that is nifty.	✓ 1.00	
<input type="radio"/> I dispute the value of this question as assessment.		
Total	1.00 / 1.00	

Question 10

OK, fine. Squint really hard and look at the top-right of the dendogram. You'll see at the very top fork, that a branch goes down the right side. How many nodes are in this branch? (e.g. how many data points end up in this branch). Note that an immediate branch going to a small subset of the data indicates strong outliers.

You entered:

245

Your Answer		Score	Explanation
245	✖	0.00	
Total		0.00 / 1.00	