Feedback — XIII. Clustering

Help

You submitted this quiz on **Fri 9 May 2014 12:56 PM IST**. You got a score of **5.00** out of **5.00**.

Question 1

For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.

Your Answer		Score	Explanation
☑ From the user usage patterns on a website, figure out what different groups of users exist.	✔	0.25	We can cluster the users with K-means to find different, distinct groups.
☑ Given a database of information about your users, automatically group them into different market segments.	✔	0.25	You can use K-means to cluster the database entries, and each cluster will correspond to a different market segment.
☐ Given historical weather records, predict if tomorrow's weather will be sunny or rainy.	✔	0.25	K-means cannot make classification predictions, as it does not label its inputs.
☐ Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.	✔	0.25	Such a prediction is a regression problem, and K-means does not use labels on the data, so it cannot perform regression.
Total		1.00 / 1.00	

Question 2

Suppose we have three cluster centroids $\mu_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$ and $\mu_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.

Furthermore, we have a training example $x^{(i)} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. After a cluster assignment step, what

will $c^{(i)}$ be?

Your Answer		Score	Explanation
$c^{(i)} = 2$			
$c^{(i)} = 3$			
$c^{(i)}$ is not assigned			
$c^{(i)} = 1$	✔	1.00	$x^{(i)}$ is closest to μ_1, so $c^{(i)} = 1$
Total		1.00 / 1.00	

Question 3

K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two?

Your Answer		Score	Explanation
☐ Using the elbow method to choose K.	✔	0.25	The choice of K must be made before running the main body of the K-means algorithm.
☐ Test on the cross-validation set.	✔	0.25	Any sort of testing is outside the scope of the K-means algorithm itself.
☑ The cluster assignment step, where the parameters $c^{(i)}$ are updated.	✔	0.25	This is the correst first step of the K-means loop.
☑ Move the cluster centroids, where the centroids μ_k are updated.	✔	0.25	The cluster update is the second step of the K-means loop.
Total		1.00 / 1.00	

Question 4

Suppose you have an unlabeled dataset $\{x^{(1)}, \ldots, x^{(m)}\}$. You run K-means with 50 different random initializations, and obtain 50 different clusterings of the data. What is the recommended

way for choosing which one of these 50 clusterings to use?

Your Answer		Score	Explanation
○ The only way to do so is if we also have labels $y^{(i)}$ for our data.			
◉ Compute the distortion function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_k)$, and pick the one that minimizes this.	✔	1.00	A lower value for the distortion function implies a better clustering, so you should choose the clustering with the smallest value for the distortion function.
○ Plot the data and the cluster centroids, and pick the clustering that gives the most "coherent" cluster centroids.			
○ Always pick the final (50th) clustering found, since by that time it is more likely to have converged to a good solution.			
Total		1.00 / 1.00	

Question 5

Which of the following statements are true? Select all that apply.

Your Answer		Score	Explanation
☐ K-Means will always give the same results regardless of the initialization of the centroids.	✔	0.25	K-means is sensitive to different initializations, which is why you should run it multiple times from different random initializations.
☑ If we are worried about K-means getting stuck in bad local optima, one way to ameliorate (reduce) this problem is if we try using multiple random initializations.	✔	0.25	Since each run of K-means is independent, multiple runs can find different optima, and some should avoid bad local optima.

☑ A good way to initialize K-means is to select K (distinct) examples from the training set and set the cluster centroids equal to these selected examples.	✔	0.25	This is the recommended method of initialization.
☐ The standard way of initializing K-means is setting $\mu_1 = \cdots = \mu_k$ to be equal to a vector of zeros.	✔	0.25	This is a poor initialization, since every centroid needs to start in a different location. Otherwise, each will be updated in the same way at each iteration and they will never spread out into different clusters.
Total		1.00 / 1.00	