

# Week 8

---

## Quiz 1

### Ques1.

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

- ☒ From the user usage patterns on a website, figure out what different groups of users exist.
- ☒ Given a set of news articles from many different news websites, find out what are the main topics covered.
- ☐ Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
- ☐ Given many emails, you want to determine if they are Spam or Non-Spam emails.

### Ques2.

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?

- ☒  $c^{(i)} = 1$
- ☐  $c^{(i)} = 2$
- ☐  $c^{(i)}$  is not assigned
- ☐  $c^{(i)} = 3$

### Answer:

- A) distance from  $[-1;2]$  to  $[1;2] = 2^2+0 = 4$
- B) distance from  $[-1;2]$  to  $[-3;0] = 2^2+2^2 = 8$
- C) distance from  $[-1;2]$  to  $[4;2] = 5^2+0 = 25$
- D) Thus, this training example will be assigned to centroid 1.

### Ques3.

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

- ☒ The cluster assignment step, where the parameters  $c^{(i)}$  are updated.
- ☐ Test on the cross-validation set.
- ☒ Move the cluster centroids, where the centroids  $\mu_k$  are updated.
- ☐ Randomly initialize the cluster centroids.

### Ques4.

Suppose you have an unlabeled dataset  $\{x^{(1)}, \dots, x^{(m)}\}$ . 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?

- ☐ Plot the data and the cluster centroids, and pick the clustering that gives the most "coherent" cluster centroids.
- ☐ Manually examine the clusterings, and pick the best one.
- ☒ Compute the distortion function  $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$ , and pick the one that minimizes this.
- ☐ Use the elbow method.

### Ques5.

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

- ☐ Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid
- ☐ K-Means will always give the same results regardless of the initialization of the centroids.
- ☐ On every iteration of K-means, the cost function  $J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_k)$  (the distortion function) should either stay the same or decrease; in particular, it should not increase.
- ☒ 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.