

# Unsupervised Learning

Quiz, 5 questions

1  
point

1.

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

- ☐ Given many emails, you want to determine if they are Spam or Non-Spam emails.
- ☐ 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.
- ☐ From the user usage patterns on a website, figure out what different groups of users exist.



1  
point

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  $x^{(i)} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ . After a cluster assignment step, what will  $c^{(i)}$  be?

- ☐  $c^{(i)} = 2$

☐  $c^{(i)}$  is not assigned

## Unsupervised Learning

☐  $c^{(i)} = 3$

Quiz, 5 questions

☐  $c^{(i)} = 1$

---

1  
point

3.

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

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

1  
point

4.

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?

- ☐ Always pick the final (50th) clustering found, since by that time it is more likely to have converged to a good solution.
  - ☐ For each of the clusterings, compute  $\frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c(i)}\|^2$ , and pick the one that minimizes this.
  - ☐ The answer is ambiguous, and there is no good way of choosing.
  - ☐ The only way to do so is if we also have labels  $y^{(i)}$  for our data.
- 

1  
point

5.

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

- ☐ For some datasets, the "right" or "correct" value of K (the number of clusters) can be ambiguous, and hard even for a human expert looking carefully at the data to decide.
- ☐ The standard way of initializing K-means is setting  $\mu_1 = \dots = \mu_k$  to be equal to a vector of zeros.
- ☐ Since K-Means is an unsupervised learning algorithm, it cannot overfit the data, and thus it is always better to have as large a number of clusters as is computationally feasible.
- ☐

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.

## Unsupervised Learning

---

Quiz, 5 questions

- ☐ I understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account. [Learn more about Coursera's Honor Code](#)

Diogo Vasconcelos

Submit Quiz

