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.
- [x] From the user usage patterns on a website, figure out what different groups of users exist.
- [ ] Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
- [x] Given a set of news articles from many different news websites, find out what are the main topics covered.

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 $x^{(i)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. After a cluster assignment step, what will $c^{(i)}$ be?

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

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?

- [x] The cluster assignment step, where the parameters $c^{(i)}$ are updated.
- [ ] Move each cluster centroid μ_k, by setting it to be equal to the closest training example $x^{(i)}$
- [x] Move the cluster centroids, where the centroids μ_k are updated.
- [ ] The cluster centroid assignment step, where each cluster centroid μ_i is assigned (by setting $c^{(i)}$) to the closest training example $x^{(i)}$.

1 point

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?

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

1 point

5. 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
- [x] On every iteration of K-means, the cost function $J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_k)$ (the distortion function) should either stay the same or decrease; in particular, it should not increase.
- [ ] K-Means will always give the same results regardless of the initialization of the centroids.
- [x] 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.

- [x] 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

Attila Földvárszky

Use the name on your government issued ID

Submit Quiz