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.

- [x] Given a database of information about your users, automatically group them into different market segments.
- [x] Given sales data from a large number of products in a supermarket, figure out which products tend to form coherent groups (say are frequently purchased together) and thus should be put on the same shelf.
- [ ] Given historical weather records, predict the amount of rainfall tomorrow (this would be a real-valued output)
- [ ] Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

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?

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

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.
- [x] Move the cluster centroids, where the centroids μ_k are updated.
- [x] 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 $\{x^{(1)}, \ldots, x^{(m)}\}$. 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?

- ( ) 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 only way to do so is if we also have labels $y^{(i)}$ for our data.
- ( ) The answer is ambiguous, and there is no good way of choosing.

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
- [ ] 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] 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.

Unsupervised Learning

Quiz, 5 questions

I, **Mark R. Lytell**, 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

Submit Quiz