Congratulations! You passed!

Grade received 100% To pass 80% or higher

Go to next item

Unsupervised Learning

Latest Submission Grade 100%

1. For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply. **1 / 1 point**

- [ ] Given historical weather records, predict if tomorrow's weather will be sunny or rainy.
- [x] From the user usage patterns on a website, figure out what different groups of users exist.

  Correct
  We can cluster the users with K-means to find different, distinct groups.

- [ ] Given many emails, you want to determine if they are Spam or Non-Spam emails.
- [x] Given a set of news articles from many different news websites, find out what are the main topics covered.

  Correct
  K-means can cluster the articles and then we can inspect them or use other methods to infer what topic each cluster represents

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? **1 / 1 point**

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

Correct
$x^{(i)}$ is closest to μ_3, so $c^{(i)} = 3$

3. K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two? **1 / 1 point**

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

  Correct
  This is the correst first step of the K-means loop.

- [x] Move the cluster centroids, where the centroids μ_k are updated.

  Correct
  The cluster update is the second step of the K-means loop.

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? **1 / 1 point**

- ○ The answer is ambiguous, and there is no good way of choosing.
- ◉ 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.
- ○ Always pick the final (50th) clustering found, since by that time it is more likely to have converged to a good solution.

Correct
This function is the distortion function. Since a lower value for the distortion function implies a better clustering, you should choose the clustering with the smallest value for the distortion function.

5. Which of the following statements are true? Select all that apply. **1 / 1 point**

- [ ] K-Means will always give the same results regardless of the initialization of the centroids.
- [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.

  Correct
  Both the cluster assignment and cluster update steps decrese the cost / distortion function, so it should never increase after an iteration of K-means.

- [ ] Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid
- [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.

  Correct
  This is the recommended method of initialization.