Unsupervised Learning

LATEST SUBMISSION GRADE

80%

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

- [x] Given a database of information about your users, automatically group them into different market segments.

  ✓ **Correct**
  You can use K-means to cluster the database entries, and each cluster will correspond to a different market segment.

- [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.

  ✓ **Correct**
  If you cluster the sales data with K-means, each cluster should correspond to coherent groups of items.

- [ ] 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.

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

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

  ✓ **Correct**
  $x^{(i)}$ is closest to μ_1, so $c^{(i)} = 1$

3. K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two? **0 / 1 point**

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

  ! **This should not be selected**
  This is not a correct description of the cluster assignment step.

- [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

- Always pick the final (50th) clustering found, since by that time it is more likely to have converged to a good solution.
- 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.
- 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.

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

- [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 decrease the cost / distortion function, so it should never increase after an iteration of K-means.

- [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.

- [ ] K-Means will always give the same results regardless of the initialization of the centroids.
- [ ] Once an example has been assigned to a particular centroid, it will never be reassigned to another different centroid