Unsupervised Learning

5 questions

1 point

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

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

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

- [ ] $c^{(i)} = 3$
- [ ] $c^{(i)} = 2$
- [ ] $c^{(i)} = 1$

- ○ $c^{(i)}$ 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?

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

---

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?

- ○ 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.
- ○ Manually examine the clusterings, and pick the best one.
- ○ Use the elbow method.

---

1 point

5.

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

- [ ] 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.
- [ ] 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 $\mu_1 = \cdots = \mu_k$ to be equal to a vector of zeros.
- [ ] 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.

- [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

Anderson Banihirwe

Submit Quiz