Close

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 historical weather records, predict if tomorrow's weather will be sunny or rainy.



Given many emails, you want to determine if they are Spam or Non-Spam emails.



Given a set of news articles from many different news websites, find out what are the main topics covered.



From the user usage patterns on a website, figure out what different groups of users exist.

1  
point

2.

Suppose we have three cluster centroids *μ*1=[12], *μ*2=[−30] and *μ*3=[42]. Furthermore, we have a training example *x*(*i*)=[−21]. After a cluster assignment step, what will *c*(*i*) be?



*c*(*i*)=1



*c*(*i*) is not assigned



*c*(*i*)=2



*c*(*i*)=3

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.



Randomly initialize the cluster centroids.



Move the cluster centroids, where the centroids *μk* are updated.



The cluster assignment step, where the parameters *c*(*i*) are updated.

1  
point

4.

Suppose you have an unlabeled dataset {*x*(1),…,*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?



Manually examine the clusterings, and pick the best one.



Compute the distortion function *J*(*c*(1),…,*c*(*m*),*μ*1,…,*μk*), and pick the one that minimizes this.



Plot the data and the cluster centroids, and pick the clustering that gives the most "coherent" cluster centroids.



Use the elbow method.

1  
point

5.

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



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.



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



On every iteration of K-means, the cost function *J*(*c*(1),…,*c*(*m*),*μ*1,…,*μk*) (the distortion function) should either stay the same or decrease; in particular, it should not increase.



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

2 questions unanswered

Submit Quiz