Course > Unit 4 Unsupervised Learning (2 weeks) > Lecture 14. Clustering 2 > 4. Computation Complexity of K-Means and K-Medoids

4. Computation Complexity of K-Means and K-Medoids

Computation Complexity of K-Means and K-Medoids

the new representative.

And what we know, this is the place

where all the action will happen,

because in K-means in this step, we were

violating both assumptions.

First of all, we were selecting the point

which may not necessarily be the member of the original set.

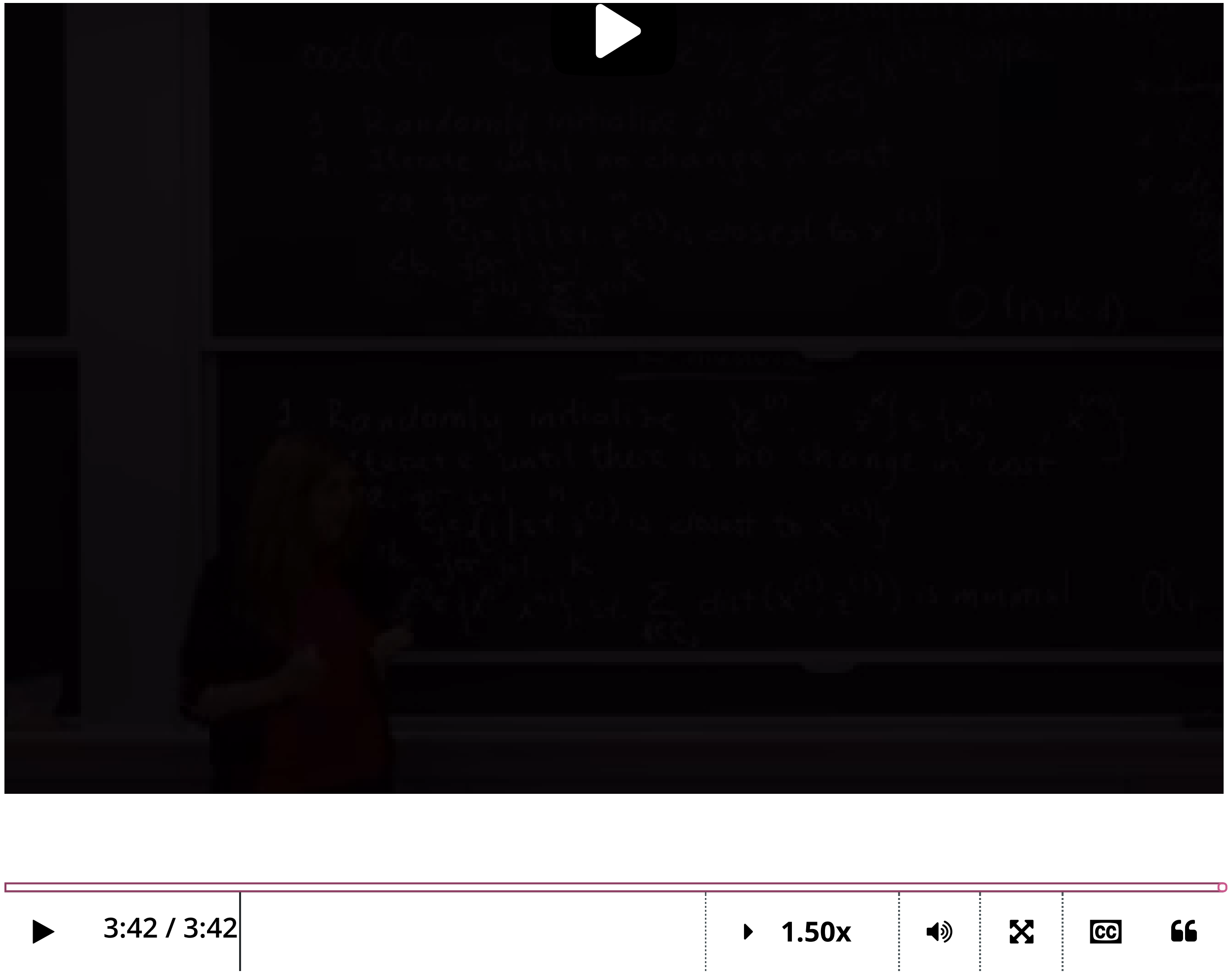
And the second point is the reason our representative look

like a centroid was because we did this derivation

and we directly built on the assumption

that our distance metric happened to be

square Euclidean distance.



So here we would expect that the change will happen.

And let's just intuitively see what we would

like to get in this step 2b.

Let's say your cluster looks like something like this.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Computation Complexity of K-Means

1/1 point (graded)

Remember that the K-Means algorithm is given by

1. Randomly select $z_1, \ldots, z_k$
2. Iterate
    1. Given $z_1, \ldots z_k$, assign each $x^{(i)}$ to the closest z_j. i.e., assign each $x^{(i)}$ such that

$$\text{Cost}\left(z_1, \ldots z_k\right) = \sum_{j=1}^{n} \min_{j=1,\ldots,k} \left\|x^{(i)} - z_j\right\|^2$$

    2. Given $C_1, \ldots, C_k$ find the best representatives $z_1, \ldots, z_k$ such that

$$z_j = \frac{\sum_{i \in C_j} x^{(i)}}{|C_j|}$$

Assuming that there are n data points $\left\{x_1, \ldots, x_n\right\}$, k clusters and representatives,and each $x_i \in \left\{x_1, \ldots, x_n\right\}$ is a vector of dimension d, what is the computational complexity for one complete iteration of the k-means algorithm?

○ $\mathcal{O}\left(n\right)$

○ $\mathcal{O}\left(nk\right)$

○ $\mathcal{O}\left(nk^2\right)$

● $\mathcal{O}\left(ndk\right)$ ✓

Solution:

In line 2.1, we go through each of the n x_i, and iterate through each of the k z_j's for each x_i (to find the closest z_j). This iteration is $\mathcal{O}\left(nk\right)$. And because each x_i has length d, the total iteration is $\mathcal{O}\left(ndk\right)$.

Line 2.2 is similar.

Note that because 2.1 and 2.2 both take $\mathcal{O}\left(ndk\right)$, one complete iteration takes $\mathcal{O}\left(ndk\right)$.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Computation Complexity of K-Medoids

2/2 points (graded)

Remember that the K-Medoids algorithm is given by

1. Randomly select $z_1, \ldots, z_k$

2. Iterate

   1. Given $z_1, \dots z_k$, assign each $x^{(i)}$ to the closest z_j. i.e., assign each $x^{(i)}$ such that

$$\text{Cost}\left(z_1, \dots z_k\right) = \sum_{j=1}^{n} \min_{j=1,\dots,k} \left\| x^{(i)} - z_j \right\|^2$$

   2. Given $C_j \in \left\{C_1, \dots, C_k\right\}$ find the best representative $z_j \in \left\{x_1, \dots, x_n\right\}$ such that

$$\sum_{x^{(i)} \in C_j} \text{dist}\left(x^{(i)}, z_j\right)$$

   is minimal.

What is the complexity of step 2a?

○ $\mathcal{O}\left(n\right)$

○ $\mathcal{O}\left(nk\right)$

○ $\mathcal{O}\left(nk^2\right)$

● $\mathcal{O}\left(ndk\right)$ ✓

Now what is the complexity of step 2b?

○ $\mathcal{O}\left(ndk\right)$

○ $\mathcal{O}\left(nk^2\right)$

○ $\mathcal{O}\left(nk^2d\right)$

● $\mathcal{O}\left(n^2dk\right)$ ✓

Solution:

Note that step 2.1 of the K-Medoids is the same as that of K-Means, so the time complexity is $\mathcal{O}\left(ndk\right)$. Note that step 2.2 of K-Medoids has an additional loop of iterating through the n points $z_j \in \left\{x_1, \ldots, x_n\right\}$which takes $\mathcal{O}\left(n\right)$. Thus step 2.2 takes $\mathcal{O}\left(n^2dk\right)$.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 14. Clustering 2 / 4. Computation Complexity of K-Means and K-Medoids

Add a Post

Show all posts ▼ by recent activity ▼

? [Staff] minor typo at Computation Complexity of K-Means
at the line 2.1 of K-Medoids algorithm, the sum index should be i, not j. 2

? [Staff] Is the .txt file transcript correct for this lecture?
It appears that the downloadable transcript for this section belongs with the previous section. 1

Learn About Verified Certificates