

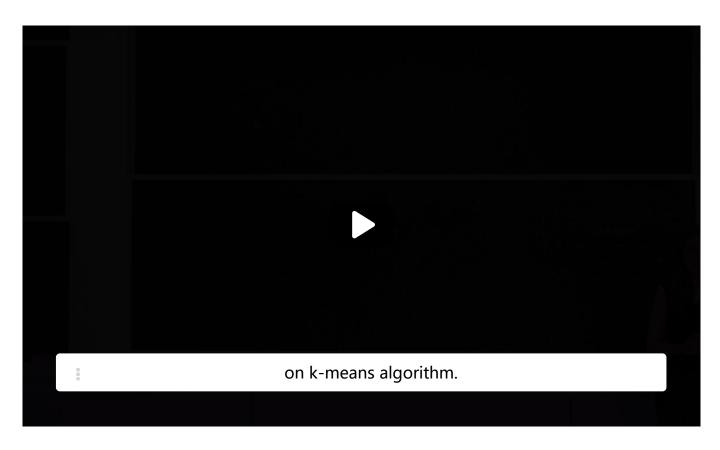
<u>Unit 4 Unsupervised Learning (2</u>

8. The K-Means Algorithm: the

<u>Course</u> > <u>weeks</u>)

> <u>Lecture 13. Clustering 1</u> > Specifics

8. The K-Means Algorithm: the Specifics The K-Means Algorithm: the Specifics



seen the k-means algorithm, we've seen the conversion properties of this algorithm,

and we also realize an important drawback of this algorithm,

that it's very sensitive to initialization.

And we understood how the properties of initialization

will be impacting the final results.

So with that, we completed the discussion

on k-means algorithm.

10:34 / 10:34

▶ 1.0x

4》

CC

End

End of transcript. Skip to the start.

Video Download video file

Transcripts

<u>Download SubRip (.srt) file</u>

<u>Download Text (.txt) file</u>



Finding the Representative Z

3/3 points (graded)

Find a simplified form of the following expression:

$$rac{\partial}{\partial z_{j}}\sum_{i\in\mathbb{C}_{j}}\left\Vert x^{(i)}-z_{j}
ight\Vert ^{2}$$

$$left \sum_{i \in \mathbb{C}_j} -2 \left(x^{(i)} - z_j
ight) oldsymbol{\checkmark}$$

$$\bigcirc \ \ -2\left(z_{j}-\sum_{i\in \mathbb{C}_{j}}x^{(i)}
ight)$$

$$igcip \sum_{i\in \mathbb{C}_j} -\left(x^{(i)}-z_j
ight)$$

$$^{\circ} \sum_{i \in \mathbb{C}_j} x^{(i)}$$

Now, what is the value of z_j that minimizes the sum?

$$egin{array}{ccc} & rac{\sum_{i \in C_j} x^{(i)}}{|C_j|}
ightarrow & \end{array}$$



Regarding update of z_j , which of the following statements is true (select all that apply)?

- lacksquare The value of z_j is affected by points $x_i
 otin C_j$
- lacksquare The value of z_j is only affected by points $x_i \in C_j$
- lacktriangledown The obtained z_j is the centroid (center of mass assuming each $x^{(i)}$ has equal mass) of the jth cluster $m{\checkmark}$



Solution:

Note that

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

is the center of mass, or centroid, of the jth cluster.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Impact of Intialization

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,\dots,z_K , assign each data point $x^{(i)}$ to the closest z_j , so that

$$\operatorname{Cost}\left(z_{1},\ldots z_{K}
ight)=\sum_{i=1}^{n}\min_{j=1,...,k}\left\Vert x^{\left(i
ight)}-z_{j}
ight\Vert ^{2}.$$

2. Given C_1,\ldots,C_K find the best representatives z_1,\ldots,z_K , i.e. find z_1,\ldots,z_K such that

$$z_j = \operatorname{argmin}_z \sum_{i \in C_j} \left\| x^{(i)} - z
ight\|^2.$$

Which of the following is true about the initialization and output of the K-Means algorithm? Select all those apply.

✓ Step 2.1 decreases or does not change the cost of clustering output ✓

✓ Step 2.2 decreases or does not change the cost of clustering output ✓	
☑ The clustering output that the K-Means algorithm converges to depends on the intialization ✓	
Solution:	
While steps 2.1 and 2.2 of the algorithm always decreases the cost or keeps it the same at least, the output of the alg dependes on the intialization of step 1. Thus, in practice, it is wise to make sure that $z_1,\dots z_K$ are intialized so that tout. Another alternative is to try multiple initializations and choose the clustering output that appears the most comp	hey are well spread
Submit You have used 2 of 3 attempts	
Answers are displayed within the problem	
What if K is 1?	
/1 point (graded) Now, assume that we are given with $K=1$ as the number of clusters. Now, does initialization matter at all?	
■ No, because cluster assignment does not change in step 2.1 ✓	
Yes, because representative selection changes in step 2.2	
Solution:	
Because if $K=1$ cluster assignment can never change, initialization does not matter. Also note that the algorithm value and same representative from there on) after just 1 iteration.	vill converge (have
Submit You have used 1 of 1 attempt	
Answers are displayed within the problem	
Discussion	Show Discussion
opic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 13. Clustering 1 / 8. The K-Means Algorithm: he Specifics	

© All Rights Reserved