edX

Audit Access Expires May 11, 2020
You lose all access to this course, including your progress, on May 11, 2020.

8. The K-Means Algorithm: The Specifics

The K-Means Algorithm: The Specifics

| ▶ | 0:00 / 0:00 | | | ▶ 1.25x | ◀)) | ⤢ | CC | 66 |

Video
Download video file

Transcripts
Download SubRip (.srt) file

Download Text (.txt) file

Finding the Representative z

3/3 points (graded)

In this problem, we will find the "best" representative z_j for the cluster $\left\{x^{(i)}\right\}_{i\in\mathbb{C}_j}$.

First, compute the following gradient:

$$\nabla_{z_j}\left(\sum_{i\in\mathbb{C}_j}\left\|x^{(i)}-z_j\right\|^2\right).$$

- ⦿ $\displaystyle\sum_{i\in\mathbb{C}_j} -2\left(x^{(i)} - z_j\right)$

- ○ $-2\left(z_j - \displaystyle\sum_{i\in\mathbb{C}_j} x^{(i)}\right)$

- ○ $\displaystyle\sum_{i\in\mathbb{C}_j} -\left(x^{(i)} - z_j\right)$

- ○ $\displaystyle\sum_{i\in\mathbb{C}_j} x^{(i)}$

✔

Find z_j that minimizes the sum $\displaystyle\sum_{i\in\mathbb{C}_j} \left\lVert x^{(i)} - z_j\right\rVert^2$.

- ⦿ $\dfrac{\sum_{i\in C_j} x^{(i)}}{\lvert C_j\rvert}$

- ○ $\displaystyle\sum_{i\in C_j} x^{(i)}$

✔

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

☐ The value of z_j is affected by points $\{x_i : i \notin C_j\}$

☑ The value of z_j is only affected by points $\{x_i : i \in C_j\}$

☑ The obtained z_j is the centroid (center of mass assuming each $x^{(i)}$ has equal mass) of the jth cluster

✔

Solution:

Note that

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

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

Submit   You have used 1 of 3 attempts

ℹ   Answers are displayed within the problem

Impact of Initialization

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 data point $x^{(i)}$ to the closest z_j, so that

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

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

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

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 initialization

✔

Solution:

While Steps 2.1 and 2.2 of the algorithm always decrease the cost or keep it the same at least, the output of the algorithm largely depends upon the initialization in Step 1. Thus, in practice it is wise to make sure that $z_1, \ldots z_K$ are initialized so that they are well spread out. Another alternative is to try multiple initializations and choose the clustering output that appears the most commonly.

[Submit]   You have used 1 of 3 attempts

ⓘ Answers are displayed within the problem

What if K is 1?

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 will converge (have same assignment and same representative from there on) after just 1 iteration.

Submit	You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

K-Means Drawbacks

0/1 point (graded)

Which of the following are drawbacks of the K-means algorithm with Euclidean distance (as presented so far in this lecture)? Select all those apply.

☑ Does not scale well to large datasets

☑ Manual choice of K ✔

☑ Not robust to outliers ✔

☑ Does not scale well with increasing number of dimensions ✔

✖

Solution:

We examine the choices in order:

Does not scale well to large datasets: This is not a disadvantage of the K-means algorithm. The complexity of the iterative algorithm scales only linearly in the number of data points.

Manual choice of K: This is a disadvantage of the K-means algorithm. The algorithm, as presented so far, does not have a built-in mechanism to choose the value of K automatically.

Not robust to outliers: This is a disadvantage of the K-means algorithm. Centroids can be dragged around by outliers or outliers might get their own cluster.

Does not scale well with increasing number of dimensions: This is a disadvantage of the K-means algorithm. With increasing number of dimensions, a distance-based similarity measure converges to a constant value between any given examples.

Submit   You have used 2 of 2 attempts

ℹ   Answers are displayed within the problem

Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 13.
Clustering 1 / 8. The K-Means Algorithm: The Specifics

<div style="text-align:right">**Hide Discussion**</div>

Add a Post

Show all posts	by recent activity

An example of interesting properties of Euclidean distance in higher dimensions

⚲ Pinned

6

Clustering K is 1

Without revealing too much I would like to discuss the case when K=1, I understand that proc...

1

Number of Clusters

Isn't choosing the number of clusters independent of what distance measure one uses?

👤 Community TA

4

The slides for lectures 13 and 14 have not been posted.

The slides for lectures 13 and 14 have not been posted. Please make them available as the pr...

2

K-means drawbacks

i found the wording of the question (so far in this lecture misleading). maybe I missed it but d...

2

Another K-means algorithm drawback

4

I Programmed The K Means Algorithm

The following is a program I made to simulate the K Means Algorithm. I used the Charlie Bro...

1

[staff] The question "What if K is 1?" is not clear

Hi Staff, the question is, "Now, does initialization matter at all?" What does it mean "to matter...

6

Representative of z

2

Meaning of "k-means does not scale well"

Is the meaning of "does not scale well" the same as "the worst case complexity (big-O) is poly...

2

How K means algorithm scale with the increasing number of dimensions?

hi I couldn't understand the reason given - [edited by staff] How K means algorithm scale wit...

2