6. Similarity Measures-Cost functions

Exercises due Apr 19, 2023 08:59 -03 Completed

Similarity Measures-Cost functions

0:00 / 0:00 1.0x

Video

Download video file

Transcripts

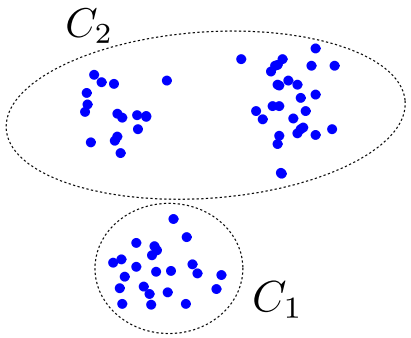
Download SubRip (.srt) file

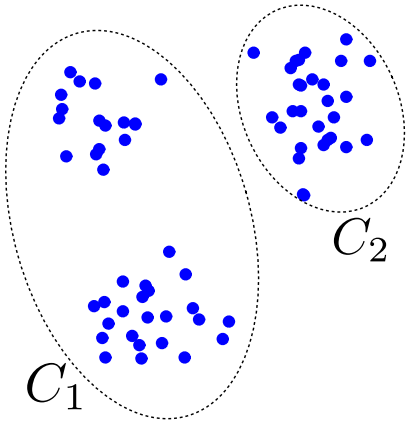
Download Text (.txt) file

The Need to Define Costs

1/1 point (graded)

Note that it is possible to have multiple clustering results given the same set of feature
the following picture, we can have two scenarios of clustering outputs given the same





Choosing the Right Similarity Measure

1/1 point (graded)

Now, let's think about the Google News example the professor has mentioned in the be
We want to measure the similarity between two Google News articles.

In the feature space, each article is represented as with the bag-of-words approach. Fo
"you", "more", "than", "Kevin" are the list of all unique vocabulary mentioned in all article
is represented as a vector while another article "you love Kevin more t
a vector . Note that each entry of vector is a binary indicator whether g
article or not.

You assume that the length of an article does not tell any useful information about the a
choose a similarity measure that does not depend on the length of the article.

Which of the following similarity measure could be the one you chose?

- ○ Euclidean distance
- ◉ Cosine distance

✔

Submit You have used 1 of 1 attempt

Diameter

1/1 point (graded)

Recall from the above video that the diameter of a cluster is the measure of how far the
cluster are located.

Choose whether the following statement is **True** or **False**: “Adding a point to a cluster e
diameter of the cluster or the diameter stays the same."

- ◉ True
- ○ False

✔

- [x] The diameter of a cluster
- [x] The average distance between points inside a cluster
- [x] The sum of distance between the representative and all points inside a cluster

✔

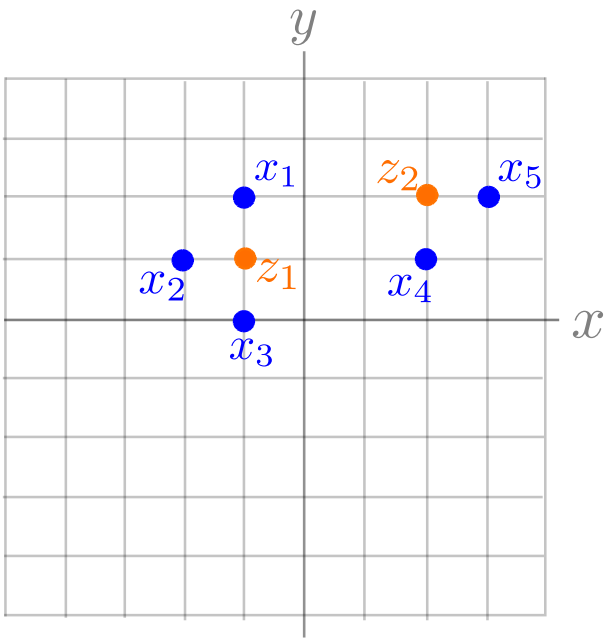
Submit You have used 1 of 3 attempts

Calculating Costs

3/3 points (graded)

As in the picture below, the set of feature vectors is given by

and the number of clusters . is clustered such that



Previous Next
3