Clustering and Similarity

Quiz, 6 questions

1 point

1.

A country, called *Simpleland*, has a language with a small vocabulary of just *“the”*, *“on”*, *“and”*, *“go”*, *“round”*, *“bus”*, and *“wheels”*. For a word count vector with indices ordered as the words appear above, what is the word count vector for a document that simply says *“the wheels on the bus go round and round.”*

Please enter the vector of counts as follows: If the counts were [*"the"*=1, *“on”*=3, *"and"*=2, *"go"*=1, *"round"*=2, "*bus"*=1, "*wheels"*=1], enter 1321211.

Enter answer here

Clustering and Similarity

Quiz, 6 questions

1 point

2.

In *Simpleland*, a reader is enjoying a document with a representation: [1 3 2 1 2 1 1]. Which of the following articles would you recommend to this reader next?

- **[7 0 2 1 0 0 1]**
- **[1 7 0 0 2 0 1]**
- **[1 0 0 0 7 1 2]**
- **[0 2 0 0 7 1 1]**

1 point

3.

A corpus in *Simpleland* has 99 articles. If you pick one article and perform **1-nearest neighbor search** to find the closest article to this query article, how many times must you compute the similarity between two articles?

- **98**
- **98*2 = 196**
- **98/2 = 49**
- **(98)^2**
- **99**

Clustering and Similarity

Quiz, 6 questions

1 point

4.

For the TF-IDF representation, does the relative importance of words in a document depend on the base of the logarithm used? For example, take the words "*bus*" and "*wheels*" in a particular document. Is the ratio between the TF-IDF values for "*bus*" and "*wheels*" different when computed using log base 2 versus log base 10?

- Yes
- No

1 point

5.

Which of the following statements are **true?** (*Check all that apply*):

- **Deciding whether an email is *spam* or *not spam* using the text of the email and some *spam / not spam* labels is a supervised learning problem.**
- **Dividing emails into two groups based on the text of each email is a supervised learning problem.**
- **If we are performing clustering, we typically assume we either do not have or do not use class labels in training the model.**

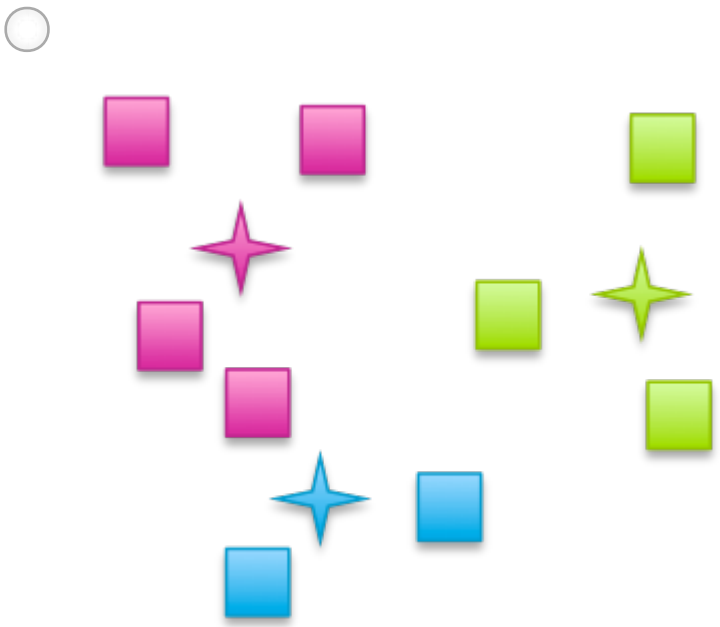
Clustering and Similarity

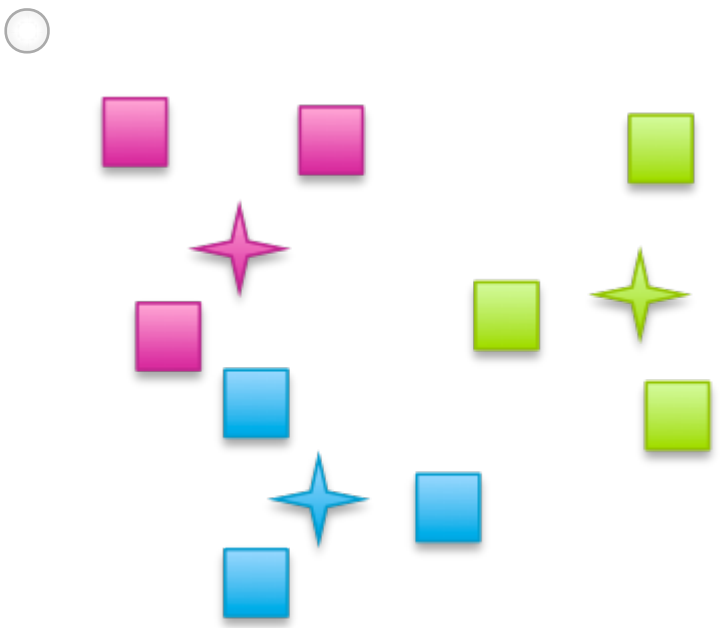
Quiz, 6 questions

1 point

6.

Which of the following pictures represents the ***best*** k-means solution? (*Squares represent observations, plus signs are cluster centers, and colors indicate assignments of observations to cluster centers.*)





I understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Clustering and Similarity

Quiz, 6 questions

Rahul kumar

Submit Quiz