


## PROBLEM 6-1 (1/1 point)

Suppose are given a stack of documents and are told that documents with similar sets of keywords are about the same topic. Your job is to organize the documents as best you can by topic. The following 4 questions refer to this situation.


For this situation, it is best to use an unsupervised learning algorithm.

- ☒ True 
- ☐ False

You have used 1 of 1 submissions

## PROBLEM 6-2 (1/1 point)


Given the above information, which of the following would be the most appropriate feature to use?

- ☐ The author's name.
- ☐ The number of pages in a document.
- ☒ The number of times a particular keyword appears in a document. 
- ☐ The number of times particular keywords and common words (for example, "the", "in", "at") appear in a document.

You have used 1 of 1 submissions

## PROBLEM 6-3 (1/1 point)


Your boss comes back with a list of 60 specific keywords as well as 5 specific topics that each keyword is best associated with. Which of the following is true, given this additional information?

- ☐ We can switch to a supervised learning algorithm.
- ☐ We can use the k-means clustering algorithm with  $k = 60$
- ☒ We can use the k-means clustering algorithm with  $k = 5$  

You have used 1 of 1 submissions

## PROBLEM 6-4 (1/1 point)

Your boss comes back one last time with new information. He can now tell you the topic of each document. However, he found some more documents for which the topic is still unknown. Given this information, can we use a supervised learning algorithm to classify the new documents?

- ☒ Yes 
- ☐ No

## PROBLEM 6-5 (1/1 point)

Remember that in Problem Set 6, we used different linkage distance measures to calculate the distances between clusters and decide which cluster a point should belong to. Consider this new method of finding linkage distances, which makes use of the linkage distance methods from the problem set:


```
def mysteryLinkageDist(self, other):
    av_dist = self.averageLinkageDist(other)
    max_dist = self.maxLinkageDist(other)
    min_dist = self.singleLinkageDist(other)
    retDist = 0.0
    if av_dist == max_dist and max_dist == min_dist:
        retDist = av_dist
    elif av_dist == max_dist:
        retDist = av_dist
    elif av_dist == min_dist:
        retDist = av_dist
    elif max_dist == min_dist:
        retDist = min_dist
    else:
        retDist = random.choice([av_dist,min_dist,max_dist])
    return retDist
```

You are given the following data points with the following feature values:

```
a 4
b 9
c 9
d 9
e 8
f 0
g 8
```


Answer the following 3 questions based on the above code.

You are asked to run the hierarchical clustering algorithm from Problem Set 6 with the `singleLinkage`, `maxLinkage`, `averageLinkage`, and `mysteryLinkage` distance metrics and asked to report the results at a cutoff of 4 clusters. The final clusters will be the same, no matter which linkage we use.

- ☒ True 
- ☐ False

## PROBLEM 6-6 (1/1 point)

Now you run the hierarchical clustering algorithm from Problem Set 6 with the `mysteryLinkage` distance metric and look at the results at a cutoff of 3 clusters. The final clusters will be:

- ☒ C0:a ||| C1:b,c,d,e,g ||| C2:f 
- ☐ C0:a,f ||| C1:b,c,d ||| C2:e,g
- ☐ C0:a,f,e,g ||| C1:b,c ||| C2:d,e
- ☐ C0:a,e,g ||| C1:b,c,d ||| C2:f

## PROBLEM 6-7 (1 point possible)

Conceptually, the `mysteryLinkage` distance metric takes a vote for each data point and decides which cluster to assign the point to based on the majority vote.

- ☒ True
- ☐ False

*You have used 1 of 1 submissions*



EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

© 2014 edX, some rights reserved.

[Terms of Service and Honor Code](#)

[Privacy Policy \(Revised 4/16/2014\)](#)

### About edX

[About](#)

[News](#)

[Contact](#)

[FAQ](#)

[edX Blog](#)


[Donate to edX](#)

[Jobs at edX](#)

### Follow Us

 [Twitter](#)

 [Facebook](#)

 [Meetup](#)

 [LinkedIn](#)

 [Google+](#)