

## CSPB 2820 - Truong - Linear Algebra with Computer Science Applications

[Dashboard](#) / [My courses](#) / [2237:CSPB 2820](#) / [18 September - 24 September](#) / [Chapter 4 Quiz](#)

**Started on** Tuesday, 19 September 2023, 9:32 PM

**State** Finished

**Completed on** Tuesday, 19 September 2023, 9:35 PM

**Time taken** 2 mins 36 secs

**Grade** Not yet graded

Question **1**

Correct

Mark 1.00 out of 1.00

Look at the first line of 4.1 page 69.

Suppose we have  $N$   $n$ -vectors ...

$N$  and  $n$  must have the same value.

Select one:

☐ True

☒ False ✓

Correct

Marks for this submission: 1.00/1.00.

Question **2**

Correct

Mark 1.00 out of 1.00

$k$  is usually bigger than  $N$

Select one:

☐ True☒ False ✓

Correct

Marks for this submission: 1.00/1.00.

Question **3**

Correct

Mark 1.00 out of 1.00

Consider page 69 4.1

$k$  is always bigger than  $n$ .

(consider all the possibilities)

Select one:

☐ True☒ False ✓

Correct

Marks for this submission: 1.00/1.00.

Question **4**

Complete

Marked out of 1.00

Why is it often relatively easy to pick  $k$  from a set of 2-vectors (vectors of length 2)?

It's mainly because when the vectors are only of length 2, then when you are deciding a value for  $k$  you are essentially deciding if you think the values belong in the same cluster or two separate clusters. There is less ambiguity for the value of  $k$  when the data set is smaller.

Question **5**

Correct

Mark 1.00 out of 1.00

Consider the definitions and the **particular set of vectors** described in the first section of 4.2

Suppose we define an additional concept B,

$$B_j = \{ x_i \mid c_i = j \}$$

What is  $B_1$  ?

- ☐ a.  $\{5\}$
- ☒ b.  $\{x_2, x_3, x_4\}$
- ☐ c.  $\{c_2, c_3, c_4\}$
- ☐ d.  $\{x_5\}$
- ☐ e.  $\{2, 3, 4\}$



Your answer is correct.

Correct

Marks for this submission: 1.00/1.00.

## Question 6

Correct

Mark 1.00 out of 1.00

Consider the kmeans function in the Chapter 4 notebook.

What is x in the argument of the function?

- ☐ a. J clust.
- ☐ b. none of these
- ☐ c. The k values.
- ☐ d. The target values of the clusters.
- ☒ e. All the vectors we will iterate through.



Your answer is correct.

Correct

Marks for this submission: 1.00/1.00.

## Question 7

Correct

Mark 1.00 out of 1.00

Consider the kmeans function.

What is k?

- ☐ a. 1000
- ☒ b. A predetermined integer.
- ☐ c. An initial guess for k that will be updated in the function.
- ☐ d. none of these
- ☐ e. a counting/iterating value that will update.



Your answer is correct.

Correct

Marks for this submission: 1.00/1.00.

Question **8**

Correct

Mark 1.00 out of 1.00

K really should be smaller than N.

(Can you explain why?)

Select one:

☒ True ✓☐ False

Correct

Marks for this submission: 1.00/1.00.

Question **9**

Correct

Mark 1.00 out of 1.00

Consider `get_n_blobs`

The default is set to 3 blobs.

Try with several different values of n.

Answer true if you have done this.

Select one:

☒ True ✓☐ False

Correct

Marks for this submission: 1.00/1.00.

Question **10**

Complete

Marked out of 1.00

Reset  $n = 3$  in the initial graph.

Now run the next kmeans blocks with  $k = 2, 3, 4$

Does the data set have 3 or 4 clusters?

Discuss here and on Piazza.

With  $k = 2$ , it is very hard to see the clusters distinctly from one another.

When  $k = 3$ , there are three clusters on the graph and when  $k = 4$  there are four clusters. This is interesting because I experimented with high values of  $k$  and it seemed to form a circle in the graph.

Question **11**

Correct

Mark 1.00 out of 1.00

The next section explores how to identify  $k$  for vectors of more than 2 dimension - i.e. we can't see what  $k$  should be.

What is interesting about the 3 data sets X, Y, and Z (Blue Dots, Yellow Grid, and Green Circle) ?

- ☐ a. They have very distinct clusters.
- ☐ b. They are completely random sets.
- ☒ c. These datasets do not have distinct clusters.
- ☐ d. none of these



Your answer is correct.

Correct

Marks for this submission: 1.00/1.00.

Question **12**

Correct

Mark 1.00 out of 1.00

The benefit of the log-log plot is that it smooths out the data to help isolate more definitive changes in the slope.  
And hence we see that the slope of all three are relatively fixed.

Select one:

☒ True ✓☐ False

Correct

Marks for this submission: 1.00/1.00.



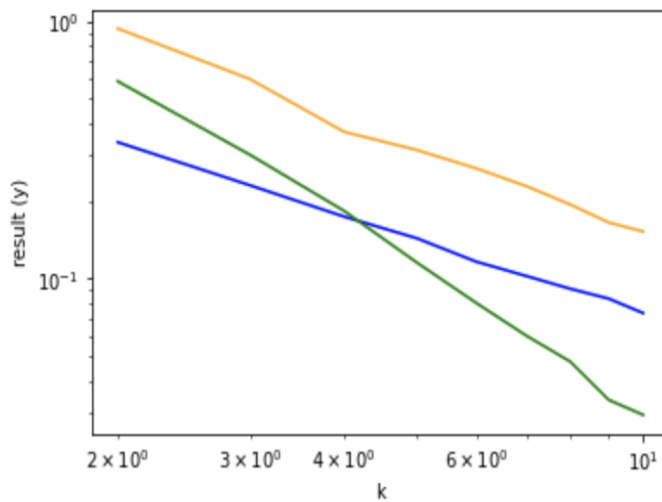
Question **13**

Correct

Mark 1.00 out of 1.00

Consider the numbers along the x-axis in the first instance of the log/log graph.

**[19]:** Text(0, 0.5, 'result (y)')



A more simple way to write these values would be:

- ☐ a. log 2, log 3, log 4, log 6, log 10 (base 8)
- ☐ b. .2, .3, .4, .6, .10
- ☐ c. Must use scientific notation to understand.
- ☒ d. 2, 3, 4, 6, 10 ,
- ☐ e. none of these.



Your answer is correct.

Correct

Marks for this submission: 1.00/1.00.

Question **14**

Correct

Mark 1.00 out of 1.00

Consider the graphs and code in the section after the words,

**"Ok, lets try it with a data set with 5 clusters"**

When we look at the line graph, we might suspect that we should use  $k = 3$  or  $k = 5$ , when clearly from the graph of the actual dots we can see 5 distinct clusters.

By looking more carefully using the **log-log plot**, it is clear that  $k = 5$ .

This suggest a technique for establishing  $k$  for sets that cannot be graphed.

Select one:

- ☒ True ✓
- ☐ False

Correct

Marks for this submission: 1.00/1.00.

Question **15**

Correct

Mark 1.00 out of 1.00

Finally, run the last couple blocks to use this method to determine  $k$  for some random datasets.

What might we guess for  $k$ ?

Discuss the challenges of finding  $k$  in Piazza.

Answer true when you have done this.

Select one:

- ☒ True ✓
- ☐ False

Correct

Marks for this submission: 1.00/1.00.

Information

Why can't we just graph the points of the last datasets to see what the best  $k$  would be?

A good thing to think about.