

Activity 1.1 - vector basics

SUBMISSION INSTRUCTIONS:

1. Complete this activity, rendering periodically to html to view your output.
2. When complete, render to pdf:
 - a. Click on the **Terminal** window (next to **Console**) and type `quarto install tinytex` then hit *Enter*
 - b. In the `yml` at the top of this document (the settings inside the `---` at the top), change the format to `format: pdf`
 - c. Render the document, make sure a pdf is created
3. Submit **both** the .pdf and the .qmd of this file to D2L.

Question 1

Consider the 3-dimensional vector $w = c(-5, 2, -3)$.

A)

Find $\|w\|_1$, the L1 (aka taxicab or Manhattan) norm of this vector. Show your work!

B)

Find $\|w\|_2$, the L2 (aka Euclidean) norm of this vector. Show your work!

C)

Create the vector in R. Use R to verify your computations above.

Question 2

Consider the `iris` data set, which comes packaged with a standard R installation:

```
data(iris)
head(iris)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

A)

Select only the sepal and petal variables. How many vectors are there, and what is the dimension of each vector?

B)

Find the mean and standard deviation vectors.

C)

Mean-center and scale the data set. Verify that the mean and standard deviation vectors of the scaled data set are $\vec{0}$ and $\vec{1}$, respectively.

D)

Find the centered/scaled vectors for the first two irises. Find the L1 and L2 distances between these two irises “from scratch,” then verify your answer using `dist`.

Question 3

Reconsider the `USairpollution` data:

```
library(HSAUR2)
```

Loading required package: tools

```
data("USairpollution")
```

A)

GOAL: Find the cities that are most and least similar with respect to their pollution. Use `dist` to find L2 distances between the cities. Then use wrangling approaches to find the two cities that are most similar, and two cities that are most dissimilar.

B)

Assess how scaling the data before computing distance impacts your answer to the previous question.

C)

Which city is Minneapolis most similar to? Most dissimilar to?