Chapter 2 Exercise 19

Exercise 19: For the following vectors, x and y, calculate the indicated similarity or distance measures.

(a) x = (1 ,1,1,1), y = (2 ,2,2,2) cosine, correlation, Euclidean

(b) x = (0 ,1,0,1), y = (1 ,0,1,0) cosine, correlation, Euclidean, Jaccard

(c) x = (0 ,−1,0,1), y = (1 ,0,−1,0) cosine, correlation, Euclidean

(d) x = (1 ,1,0,1,0,1), y = (1 ,1,1,0,0,1) cosine, correlation, Jaccard

(e) x = (2 ,−1,0,2,0,−3), y =(−1,1,−1,0,0,−1) cosine, correlation

Download a .zip file from http://webpages.uncc.edu/aatzache/ITCS6190/Exercises/Exercise19\_Chapter02\_SimilaritiesUsingVectors.zip

Unzip the downloaded file

**Steps to execute a Spark program for Similarity Measures:**

1. Open Eclipse (with Spark)

2. Create simple Scala project

3. Copy the Similarities.scala from the unzipped folder

4. Run Similarities.scala

5. When it asks for input data (use the data from above – Chapter 2 Exercise 19)

Sample input:

Enter vector size (if input vector is x=(1,1,1,1), then size is 4):

4

Enter Input vector 1 (each value innew line):

1

1

1

1

Enter Input vector 2:

2

2

2

2

6. Copy the output into a similaritymeasures.txt file (The file should contain the output for all the input vectors)

7. Take a screenshot of the scala runtime environment

8. Submit the similaritymeasures.txt and the screenshot from step 7 to canvas.