DATA-101 IN-CLASS ASSIGNMENT SEPTEMBER 23, 2021

1. (3 points) Which of the following option is true about k-NN algorithm?
   (a) It can be used for classification.
   (b) It can be used for regression.
   (c) It can be used in both classification and regression.
   (d) None of the above

2. (3 points) Which of the following statement is true about k-NN algorithm?
   (a) k-NN performs much better if all of the data have the same scale.
   (b) k-NN works well with a small number of input variables, but struggles when the number of inputs is very large.
   (c) k-NN makes no assumptions about the distribution of the data.
   (d) All of the above.
   (e) None of the above.

3. Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable.

x	y	Class
-1	1	1.1
0	1	1.7
0	2	2.4
1	-1	1.2
1	0	1.9
1	2	3.0
2	2	3.1
2	3	3.2

   (a) (4 points) Consider the point A = (0,0). Using the two-nearest neighbors, what is the predicted value for Class of the point A?
   (b) (4 points) Consider the point B = (3,3). Using the three-nearest neighbors, what is the predicted value for Class of the point B?

4. Consider the point A = (1, 3, 5, 7, 13). In Python, answer the following:
   (a) (3 points) Rescale the point A using the 0-1 transformation.
   (b) (3 points) Rescale the point A using the z-score standardization.

5. Consider the point B = (3.1, -1, 4.3, 8.2, 7.2, 2.2). In R, answer the following:
   (a) (3 points) Rescale the point B using the 0-1 transformation.
   (b) (3 points) Rescale the point B using the z-score standardization.