Assignment 1

- Individual Assignment

    a) Normalise the data using the Z-score normalisation

    By using the Z-score normalisation formula: (X – mean)/std

    → Sepal Length: mean 5.793 / standard deviation 0.887

    → Sepal Width: mean 3.107 / standard deviation 0.341

    By plugging in the values in the normalisation formula, the result is as below:

Index	Sepal_length		Sepal_width		Class
	Origin	Z-Score Normalised	Origin	Z-Score Normalised	
1	6.300	0.571	2.900	-0.605	Virginica
2	5.100	-0.781	3.400	0.859	Setosa
3	5.700	-0.105	2.500	-1.776	Virginica
4	5.000	-0.894	3.500	1.152	Setosa
5	4.800	-1.120	3.400	0.859	Setosa
6	6.600	0.909	2.900	-0.605	Versicolor
7	6.300	0.571	3.300	0.566	Versicolor
8	6.300	0.571	3.400	0.859	Versicolor
9	6.000	0.233	3.400	0.859	Versicolor
10	4.700	-1.232	3.200	0.273	??
11	5.000	-0.894	3.300	0.566	??
12	6.100	0.346	2.900	-0.605	??
13	6.800	1.135	3.200	0.273	??
14	4.500	-1.458	2.300	-2.362	??
15	7.700	2.149	3.000	-0.312	??

    b) Use the K-nearest neighbours method with k = 3 to predict the missing classes. Use the Euclidean norm in your distance calculation

    → By using the 2-norm distance, the labels of 10 ~ 15 data points are as follows.

    10: Setosa, 11: Setosa, 12; Versicolor, 13; Versicolor, 14: Virginica,

    15: Versicolor

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
10	4.700	-1.232	3.200	0.273	Setosa	
5	4.800	-1.120	3.400	0.859	Setosa	0.596
2	5.100	-0.781	3.400	0.859	Setosa	0.739
4	5.000	-0.894	3.500	1.152	Setosa	0.941
9	6.000	0.233	3.400	0.859	Versicolor	1.578
7	6.300	0.571	3.300	0.566	Versicolor	1.827
8	6.300	0.571	3.400	0.859	Versicolor	1.896
1	6.300	0.571	2.900	-0.605	Virginica	2.006
6	6.600	0.909	2.900	-0.605	Versicolor	2.315
3	5.700	-0.105	2.500	-1.776	Virginica	2.339

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
11	5.000	-0.894	3.300	0.566	Setosa	
2	5.100	-0.781	3.400	0.859	Setosa	0.314
5	4.800	-1.120	3.400	0.859	Setosa	0.370
4	5.000	-0.894	3.500	1.152	Setosa	0.586
9	6.000	0.233	3.400	0.859	Versicolor	1.164
7	6.300	0.571	3.300	0.566	Versicolor	1.465
8	6.300	0.571	3.400	0.859	Versicolor	1.494
1	6.300	0.571	2.900	-0.605	Virginica	1.876
6	6.600	0.909	2.900	-0.605	Versicolor	2.150
3	5.700	-0.105	2.500	-1.776	Virginica	2.472

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
12	6.100	0.346	2.900	-0.605	Versicolor	
1	6.300	0.571	2.900	-0.605	Virginica	0.225
6	6.600	0.909	2.900	-0.605	Versicolor	0.564
7	6.300	0.571	3.300	0.566	Versicolor	1.193
3	5.700	-0.105	2.500	-1.776	Virginica	1.255
9	6.000	0.233	3.400	0.859	Versicolor	1.468
8	6.300	0.571	3.400	0.859	Versicolor	1.481
2	5.100	-0.781	3.400	0.859	Setosa	1.848
5	4.800	-1.120	3.400	0.859	Setosa	2.071
4	5.000	-0.894	3.500	1.152	Setosa	2.150

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
13	6.800	1.135	3.200	0.273	Versicolor	
7	6.300	0.571	3.300	0.566	Versicolor	0.635
8	6.300	0.571	3.400	0.859	Versicolor	0.813
6	6.600	0.909	2.900	-0.605	Versicolor	0.907
1	6.300	0.571	2.900	-0.605	Virginica	1.044
9	6.000	0.233	3.400	0.859	Versicolor	1.075
2	5.100	-0.781	3.400	0.859	Setosa	2.003
4	5.000	-0.894	3.500	1.152	Setosa	2.211
5	4.800	-1.120	3.400	0.859	Setosa	2.329
3	5.700	-0.105	2.500	-1.776	Virginica	2.396

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
14	4.500	-1.458	2.300	-2.362	Virginica	
3	5.700	-0.105	2.500	-1.776	Virginica	1.474
1	6.300	0.571	2.900	-0.605	Virginica	2.684
6	6.600	0.909	2.900	-0.605	Versicolor	2.948
5	4.800	-1.120	3.400	0.859	Setosa	3.239
2	5.100	-0.781	3.400	0.859	Setosa	3.291
4	5.000	-0.894	3.500	1.152	Setosa	3.559
7	6.300	0.571	3.300	0.566	Versicolor	3.562
9	6.000	0.233	3.400	0.859	Versicolor	3.638
8	6.300	0.571	3.400	0.859	Versicolor	3.807

Index	Sepal_length Origin	Sepal_length Z-Score Normalised	Sepal_width Origin	Sepal_width Z-Score Normalised	Class	Distance
15	7.700	2.149	3.000	-0.312	Versicolor	
6	6.600	0.909	2.900	-0.605	Versicolor	1.274
1	6.300	0.571	2.900	-0.605	Virginica	1.605
7	6.300	0.571	3.300	0.566	Versicolor	1.806
8	6.300	0.571	3.400	0.859	Versicolor	1.965
9	6.000	0.233	3.400	0.859	Versicolor	2.246
3	5.700	-0.105	2.500	-1.776	Virginica	2.688
2	5.100	-0.781	3.400	0.859	Setosa	3.156
4	5.000	-0.894	3.500	1.152	Setosa	3.377
5	4.800	-1.120	3.400	0.859	Setosa	3.472