task_9.4

Machine Learning (WiSe 2025/2026)

Author: Suvansh Shukla
Matriculation No. 256245

Assignment 9 Task 4

Part A

Our task is to use KNN algorithm to predict the class of point P.

These are the given points, along with the newest point:

Letter	X	Y	Z	Class
A	1	1	1	1
B	9	1	2	1
C	4	2	1	1
D	6	5	4	2
E	3	4	3	2
F	1	4	4	2
P	20	4	4	?

We need to use standard Euclidean distance as our similarity function.
Formula for Euclidean distance:

$$d(x, y) := \sqrt{\sum_i (a_i(x) - a_i(y))^2}$$

Calculating distance of P to all other points (Without Normalization)

$$d(P, A) = \sqrt{(X_P - X_A)^2 + (Y_P - Y_A)^2 + (Z_P - Z_A)^2} = 19.4679$$

$$d(P, B) = \sqrt{(X_P - X_B)^2 + (Y_P - Y_B)^2 + (Z_P - Z_B)^2} = 11.5758$$

$$d(P, C) = \sqrt{(X_P - X_C)^2 + (Y_P - Y_C)^2 + (Z_P - Z_C)^2} = 16.4012$$

$$d(P, D) = \sqrt{(X_P - X_D)^2 + (Y_P - Y_D)^2 + (Z_P - Z_D)^2} = 14.0356$$

$$d(P, E) = \sqrt{(X_P - X_E)^2 + (Y_P - Y_E)^2 + (Z_P - Z_E)^2} = 17.0293$$

$$d(P, F) = \sqrt{(X_P - X_F)^2 + (Y_P - Y_F)^2 + (Z_P - Z_F)^2} = 19.0000$$

The nearest point to P is B, and rest in order of closeness are:

B, D, C, E, F, A

Assuming K=1, since B is the nearest neighbour in this case, P would be assigned class 1.

Assuming K=3, since B and C are class 1, while only D is class 2. P would be assigned class 1.

Assuming K=5, B & C are class 1, while D, E & F are all class 2. P would be assigned class 2.

Calculation after Normalization

We use the following formula for Normalization:

$$x = x - x_i x_{ax} - x_i$$

So, the X column values would change to:

Letter	X	X	Y	Y	Z	Z	Class
A	1	0	1	0	1	0	1
B	9	0.4210	1	0	2	0.3333	1
C	4	0.1578	2	0.25	1	0	1
D	6	0.2613	5	1.00	4	1	2
E	3	0.1052	4	0.75	3	0.6666	2
F	1	0	4	0.75	4	1	2
P	20	1	4	0.75	4	1	?

Re-calculating all distances:

$$d(P, A) = \sqrt{(X_P - X_A)^2 + (Y_P - Y_A)^2 + (Z_P - Z_A)^2} = 1.6007$$

$$d(P, B) = \sqrt{(X_P - X_B)^2 + (Y_P - Y_B)^2 + (Z_P - Z_B)^2} = 1.1585$$

$$d(P, C) = \sqrt{(X_P - X_C)^2 + (Y_P - Y_C)^2 + (Z_P - Z_C)^2} = 1.3997$$

$$d(P, D) = \sqrt{(X_P - X_D)^2 + (Y_P - Y_D)^2 + (Z_P - Z_D)^2} = 0.7798$$

$$d(P, E) = \sqrt{(X_P - X_E)^2 + (Y_P - Y_E)^2 + (Z_P - Z_E)^2} = 0.9548$$

$$d(P, F) = \sqrt{(X_P - X_F)^2 + (Y_P - Y_F)^2 + (Z_P - Z_F)^2} = 1.0000$$

The nearest point to P is D, and all the points in order of closeness are:

D, E, F, B, C, A

Assuming K=1, since D is the nearest neighbour in this case, P would be assigned class 2.

Assuming K=3, since D, E & F are all class 2, and none are class 1. P would be assigned class 2.

Assuming K=5, B & C are class 1, while D, E & F are all class 2. P would be assigned class 2.

After Normalization, we can see that point is far more likely to be predicted to have class 2 than class 1.

Part B

In this scenario, we are supposed to do leave-one-out-crossvalidation to predict class labels for each instance. We start out by initializing all weights to 1. All weight adjustments are in increment/decrements of 0.5. This crossvalidation strategy is done for a total of 7 times (equal to the number of instances), with one final iteration for giving the final prediction of point P, using the finalized weights from the past 7 iterations.

Note: I will use normalized values for distance calculations and weighted Euclidean distance.

Calculating all attribute-wise distances for iteration 1, i.e. predicting class for point A:

X distances (A)	Y distances (B)	Z distances (C)	total $\overline{A^2+B^2+C^2}$

|