Lab 5

KNN classification model

⇒ Consider the following dataset, for k = 3
and test data (x, 35, 100) as (Person,
Age, Salary K) value using KNN classifier
model and predict the target.

Person	Age	Salary K	Target
A	18	50	N
B	23	55	N
C	24	70	N
D	41	60	Y
E	43	70	Y
F	38	40	Y
X	35	100	?

⇒ Making of distance table

Age	Salary K	Target	Distance
18	50	N	52.81
23	55	N	46.57
24	70	N	31.953
41	60	Y	40.44
43	70	Y	31.048
38	40	Y	60.055
35	100		

Closest neighbors → E, C, D

Classes ⇒ [Y, N, Y]

⇒ majority = Y, so (X, 35, 100) belongs

For the dataset

how to choose the K value? Demonstrate using accuracy rate & error rate.

Steps are

i) Split dataset into Training and Testing set.

ii) Train K-NN with different K values like 1, 3, 5, 7 ...

iii) Calculate accuracy

$$Accuracy = \frac{Correct \ predictions}{Total \ prediction} \times 100$$

iv) Calculate error rate ⇒ Error rate = 1 - accuracy rate

v) Plot accuracy of K

For diabetes dataset

What is feature scaling and purpose of it? How to perform it??

➡ Purpose =>

i) Diabetes dataset has features like glucose, BMI, age, which have different ranges.

ii) Machine learning algos perform better when features are on similar scale.

=> It helps increase convergence speed & prevents bias from dominant features.
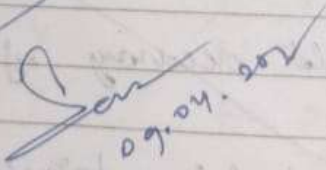
To perform, we have two ways =>

i) Min-max scaling =>

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

scales b/w 0 to 1

2) Standardization

Converts to mean 0 & standard deviation 1

$$x' = \frac{x - \mu}{\sigma}$$

09.04.204