Date 7/11/25
Page

Lab - 06

To Build a KNN classification model for the given datasets.

1. Consider the following dataset, for K=3 & test data (X, 35, 100) as (Person, Age, SalaryK) solve using KNN classifier model & predict the target.

Person	Age	SalaryK	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	?

Sol. Distance $= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$

A(18, 50) $= \sqrt{(35-18)^2 + (100-50)^2} = 52.81$

B(23, 55) $= \sqrt{(35-23)^2 + (100-55)^2} = 46.59$

C(24, 70) $= \sqrt{(35-24)^2 + (100-70)^2} = 31.95$

D(41, 60) $= \sqrt{(35-41)^2 + (100-60)^2} = 40.45$

E(43, 70) $= \sqrt{(35-43)^2 + (100-70)^2} = 31.06$

F(38, 40) $= \sqrt{(35-38)^2 + (100-40)^2} = 60.08$

Sort by Distance

E(43, 70) = 31.06 → Y

C(24, 70) = 31.95 → N

D(41, 60) = 40.45 → Y

Majority is Y

∴ (X, 35, 100) = Y

Date 7/4/25
Page

(2). Consider Iris dataset to build KNN classifier

→ How to choose the K value?

The value of K in KNN is choosen by considering:

* Accuracy Rate: Proportion of correct prediction [Higher accuracy is better]
* Error Rate: Proportion of incorrect prediction [lower rate is better]

* Tested K-values from 1 to 20 and calculated accuracy & error rate for each on the test set and choose the optimal K value

Example:

K=1; Accuracy = 0.9667, Error = 0.0333
K=3; Accuracy = 0.9667, Error = 0.0333
K=5; Accuracy = 1.0000, Error = 0.0000
K=15; Accuracy = 0.9667, Error = 0.0333

∴ K=5 has been choosen.

(3). Consider Diabetes dataset to build KNN classifier

→ What is the purpose of feature scaling? How to perform it?

Raw 7/4/25

* KNN relies on euclidean distance b/w datapoints without scaling, features with larger ranges dominate over those with smaller ranges
* Feature scaling is used to improve model performance by treating all features equally.