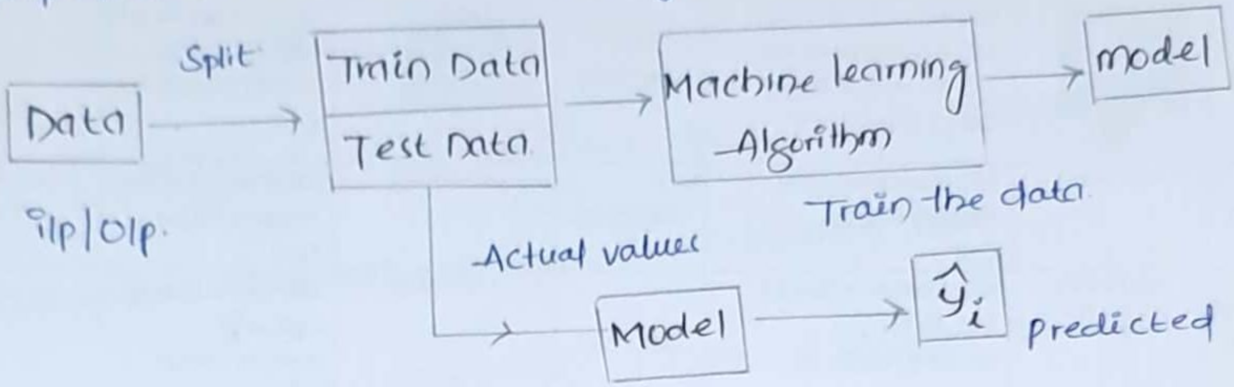
K – Nearest Neighbour (K.N.N.) :-

Before going to any algorithm we have to Understand the pipe line of machine Learning ~~alg~~ process.



i) let Consider a data Contiining input and output Generally we consider input data as (x), and output as (y) if our output label Contain Numerical (Continious) form we do regression task, if our Ouput label Contain Descrete (Categorical) we do classific-ation task. This two tasks are Comes under Super Voiced Machine Learning Techniques. most of our algorithms perform both classification and regression tasks.

ii) Similarly K.N.N. algorithm also perform both the tasks, then after we have to Split data as train $-test then apply Some machine Learning algorithm to train data. (ie Train the data Intially).

iii) ~~After~~ it we are making a model. then take X_{test} data points and deploying in model it makes y_{test_pred} values.

Then Compare finally with altual valuee.

K.N.N :- it is a Supervised Learning technique algorithm, which is a Simple algorithm, Solve both regression and classification task.

Now we discuss with classification task in K.N.N.

classification :- If we have data contain output label descreate or categorical then we perform classification task.

Let Consider data containing Heights & weights as input variable and gender is output variable, here Our task is to predict the gender which is Categorical type by using "K.N.N." classifier

Intially take points and plot a graph in between heigh and weight On x-axis, y-axis respectively. from trained data.

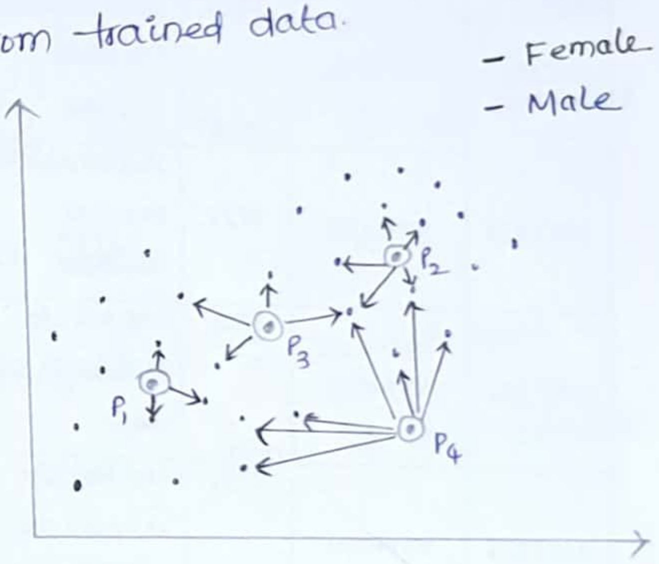
I/P Train Data O/P

H	W	G
1	3	M.
2	2	F.
3	1	M

$\xrightarrow[\text{Scatterplot}]{\text{Visualization}}$

Test Data

H	W	Gender
4	7	?
5	6	



P(1) => K = 3 => Female class

P(2) => K = 5 => Male class

P(3) => K = 4 => can't Defined

P(4) = K = 7 => Male class.

from table (H & W) are i/p variables where Gender is Output variable

So here Our task is to predict the new query points Gender as per train data with KNN classifier algorithm

Lets Consider Query points as x_q in that we want to find its class as shown in figure by Geometrical way ie x_q is nothing but new points (P_1, P_2, P_3, P_4) by Using k-value ie nearest neighbours if we observe this Containing distances. we know if we have find distances we use "Eucledian Distances".

Let Consider point (P_1) we Consider k-value as '3' ie take nearest neighbour points from that point only Consider '3' points as shown in figure. So (P_1) represent all of class belonging to "Female".

Then Consider point (P_2) ⇒ $k=5$ ⇒ ie Consider '5' nearest neighbour points from (P_2) that shows all are from class "male". So we Consider P_2 as male class and (P_1) as female class.

Similarly Consider point (P_3) ⇒ $k=4$ ⇒ P_3 shows 2 nearest points as female and remaining 2 nearest points as male in that Situation we didn't Say anything about this (P_3) class.

Note :- In K.N.N. Generally we consider majority of points if mejority of points belonging to "Male" we Conclude the x_q belonging to male for this process Generally we Consider "odd" numbers are k-value if we Consider 'Even" Numbers then we Can't Determind.

Let Consider point (P_4) ⇒ $k = 7$ ⇒ shows 4 male points and 3 female points by majority of nearest point we say (P_4), is male class.

Pseudo Code for k-NN Algorithm :-

Given → Data ⟨ i/p, o/p → categorical → classification.

Task → x_q → $\hat{y}_q$ Predicted value

Algorithm :- K.N.N. (k-value = ?)

i) Intialize the distances in list format

ii) for each x_i in the data set - Compute the distances

ie $d(x_i, x_q) \rightarrow d_i$ ⇒ distance b/w point to Query point

then we have to Store all the distances di in the list.

iii) Then Sort the list on basis of distances ie (we find the distance from query point to every point, after it we have to Sort in ascending Order every data point distance.) $[x_i, y_i, d_i]$

~~iv)~~ Ex - We have distance like [5, 6, 2, 3, 1, 7, 9, 4, 0, 1] Units

Sort → $[\underline{0, 1, 1}, 2, 3, 4, 5, 6, 7, 9]$ Units.

iv) After Sort Out as per Our k-value Consider nearest points ie if $k = 3$ then Consider [0, 1, 1] as Shortest points from x_q if that belonging to female class then we Say our Query point is female, if thatbelonging to male then we Say it male class.

v) As per k value consider nearest points and store them in a new list for suppose consider K.N.N = [0, 1, 1]

Every point consider x_i, y_i value also $[x_i, y_i, d_i]$ ↳ 0, 1, 1 → KNN

```
vi) for each x_i in K.N.N :
        if y_i == Male :
            Count_male += 1
        else :
            Count_female += 1
```

```
vii) finally we can say if count_male > count_female :
                                print ("Male class")
                        else :
                                print ("Female class")
```

Note :- This algorithm is very simple but it consist most of time ie high time complexity thats why we say K.N.N. is "Lazy Larner" Algorithm. if we have most of data points lyining in data set then we cant use K.N.N. Algorithm.