6/3/23

KNN (K-NEAREST NEIGHBOUR)

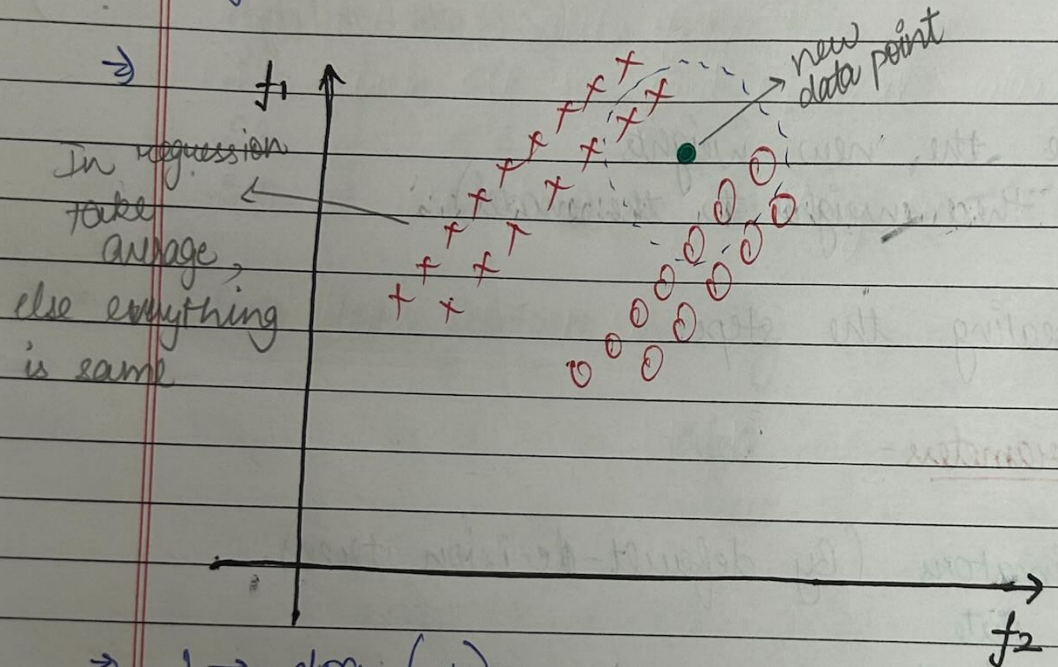
→ KNN

Used for

Regression

Classification

→



f_1	f_2	O/P
-	-	-
-	-	-

→ 1 → dog (x)
0 → cat (O)

→ In classification, categories are fixed.

→ (K)NN

↓

Hyperparameter

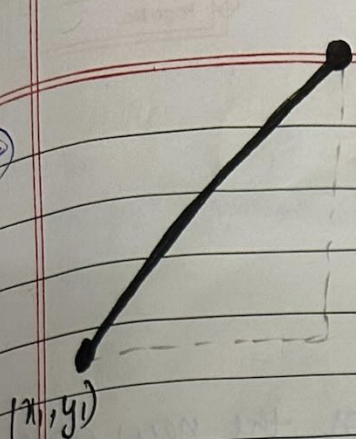
K → Grid search CV
↳ Manual

→ K = no. of nearest data points

→ if K=5, 5 nearest data points

K=3, 3 nearest data points

Date / /
Page No.



⇒ Distance of nearest points
↳ Euclidean Distance
↳ Manhattan Distance

Euclidean distance = $\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

⇒ If there are 100 data points, calculate distance of all 100 points & then select 5 nearest points.

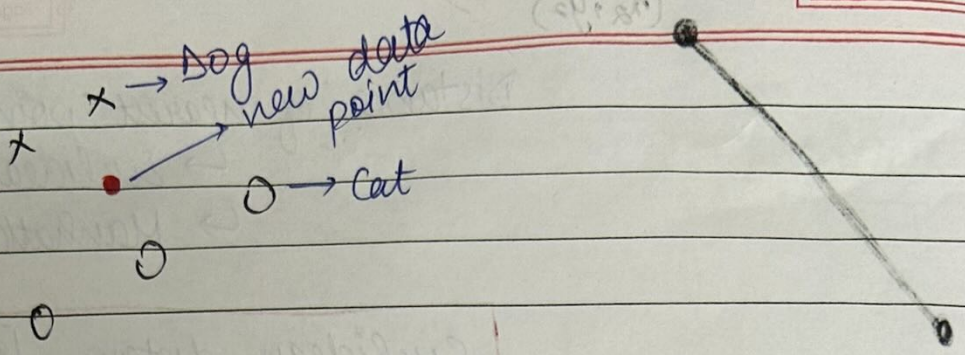
⇒ Always take odd values of K = 1, 3, 5, 7, ——
(∵ prediction is done on the basis of majority voting,
∴ in odd values, voting is done easily)

⇒ Steps -

- Step 1 - Decide the value of K
- Step 2 - Find K nearest data points
- Step 3 - To which class new point belongs (done by majority voting)

⇒ Steps -

- Step 1: Decide the value of K
- Step 2: Find the distance using euclidean distance
- Step 3: Find K nearest points.
- Step 4: Select the class to which the point belong (done through majority voting)



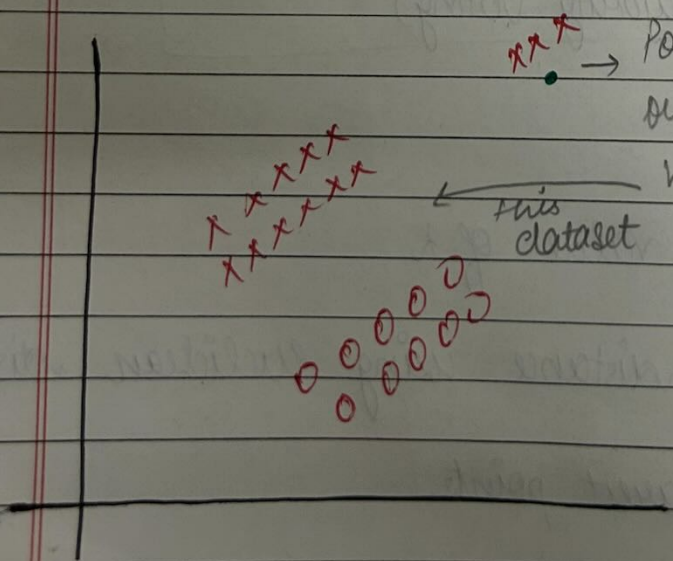
Use voting for finding the class for the new data point.

⇒ For Grid search CV,
Make a list of values of K = 1, 3, 5, 7 --- & use Grid Search CV to find optimal value of K.

⇒ KNN is also known as Lazy Runner

⇒ **Limitations of KNN**

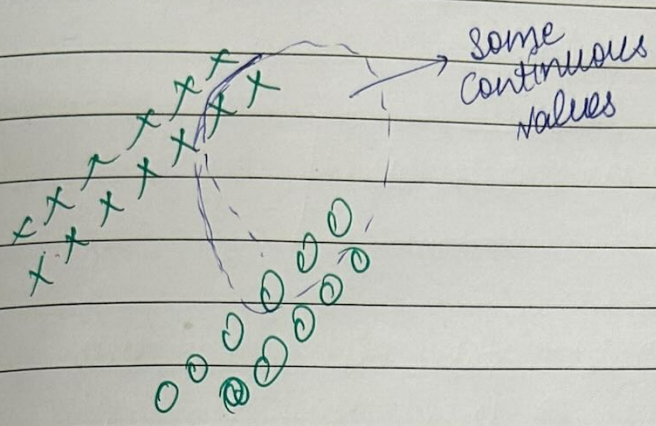
- doesn't work on huge datasets
- sensitive to outliers
- sensitive to missing values



⇒ Agar missing values ho toh model overfit ho jaata hai.

Date / /
Page No.

KNN for Regression



⇒ Take mean of all the values or average of k nearest values & the mean or average is the predicted value.

⇒ In regression take average on output.

⇒ K = 1 model hmesha overfit hota hai

⇒ Jaise k ki values increase krenge vaise vaise model acha hoga, K = 1, 3, 5 ---

⇒ Not possible for non-linear data.

* Classification → Confusion
  Precision
  Recall
  Regression → MSE & MAE
  } Evaluation metrics (of KNN, infact all)