Binary Classification

0 | 1

< 0.5 = 0

> 0.5 = 1

[0.01 , 0.99]

[0.05 , 0.95]

[0.1 , 0.9]

[0.4 , 0.6]

[0.6 , 0.4]

Probability of 0 | Probability 1

N	Y
0	1

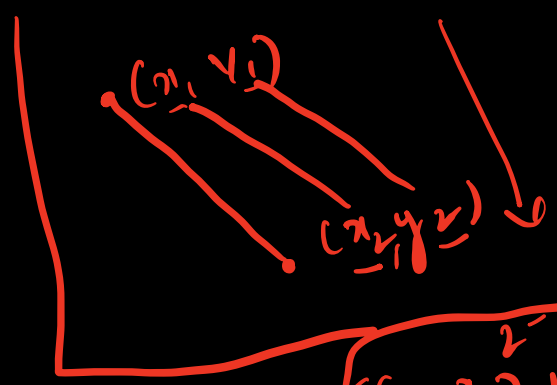
K-NN

K - Nearest Neighbour

How many Neighbours to consider? [elbow method]

Distance

- Hamming distance
- Minkowski
- Manhattan
- Euclidean



$$\left(\sum (x_1-x_2)^q + (y_1-y_2)^q\right)^{1/q}$$

$$\sum |x_1-x_2| + |y_1-y_2|$$

Linear Reg
Logistic Reg
$\Rightarrow$ o/p

model
$y = mx + b$
↑ predict
Unseen Data

K-NN $\Rightarrow$ o/p

No model
lazy model

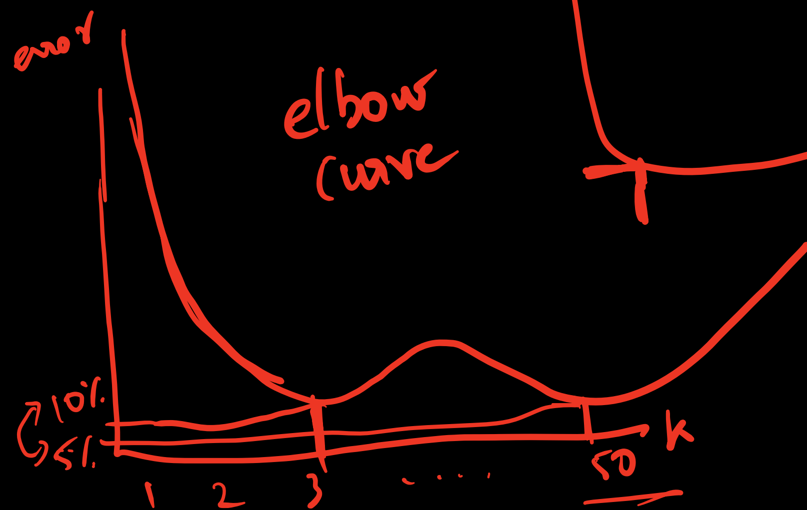
Huge Dataset
Huge performance overhead

↕ → distance
select Neig mode, mean to predict
Unseen data

Large Samples (N): Large storage
row
space complexity
time complexity

Large columns
Curse of dimensionality: Lower Accuracy

Large K : underfitting
high bias, smooth boundary
, low bias



k less

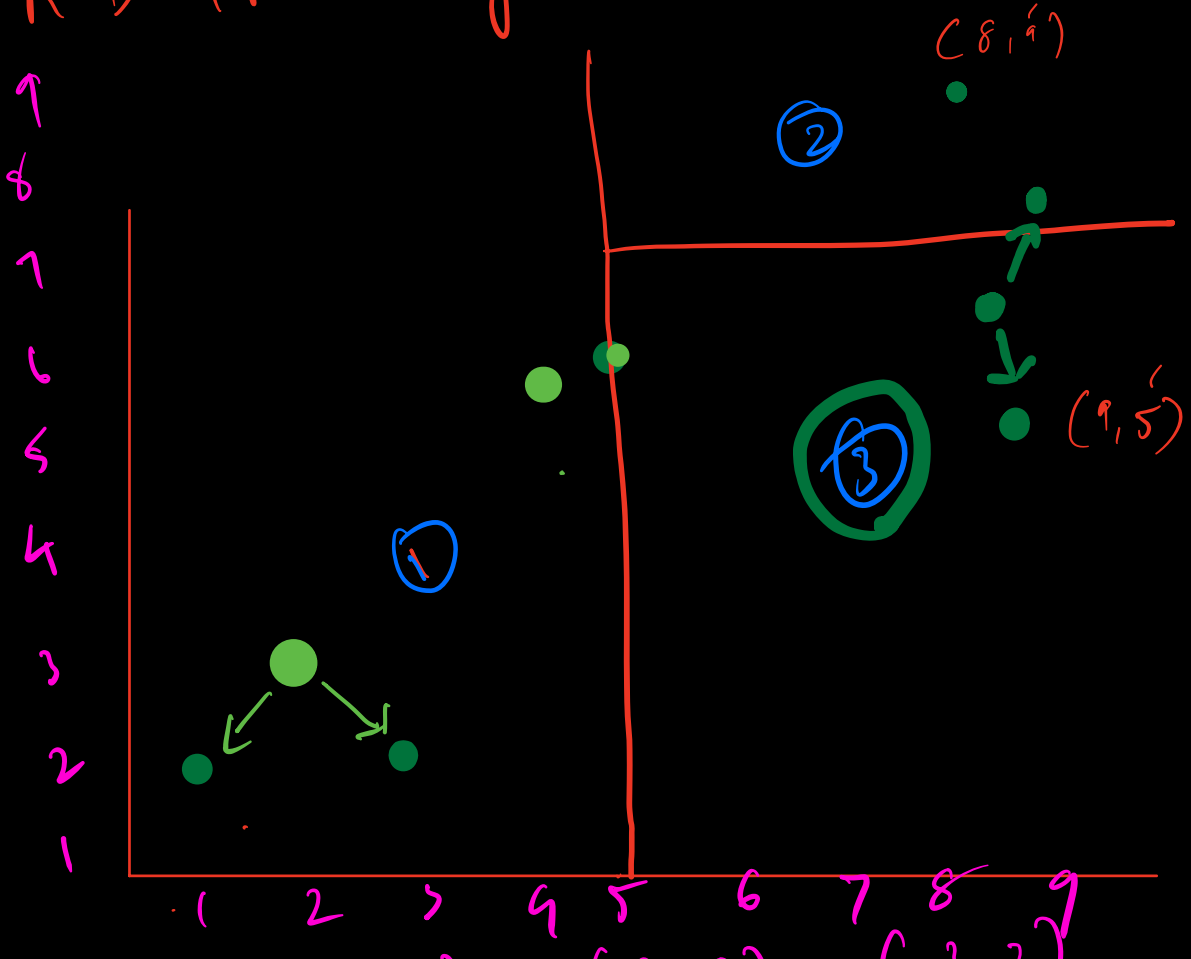
samples

1
50 | data record

50

1 millions 8 di 10k

K-D Tree Algorithm



(1,2) (5,6) (8,9) (3,2) (9,5)

1 5 8 3 9

1 3 5 8 9