Campus: PCE. Course: BTECH | Class/Section: VI CSE-A | Date: 24-03-21

Name of Faculty: Praveen Kumar Yadav | Name of Subject: ML | Code: 6CS6-01

KNN CLASSIFICATION Algorithm:-

↳ K Nearest Neighbours Algorithm (Compute K-Nearest neighbour and Take Majority vote).

x : +ve data point

* : -ve data point

$$D = \{(x_i, y_i) \mid x_i \in R^2, \; y_i \in \{0, 1\}\}$$

$D \rightarrow ML \rightarrow f$

$x_q \rightarrow f \rightarrow$ +ve / -ve

↳ Take all points that are nearest to x_q (Geometric close/Geometric proximity)

objective $x_q \rightarrow y_q$

↳ conclude $x_q \rightarrow$ blue.

Given TASK: give x_q, Find $x_q \rightarrow y_q \in \{0, 1\}$ (-ve, +ve)

① Find K-nearest point to x_q in D $\rightarrow$ 3 NN to x_q

Let K = 3 $\{x_1, x_2, x_3\}$

$x_1 \downarrow y_1$, $x_2 \downarrow y_2$, $x_3 \downarrow y_3$

② Take their class-Labels $\{y_1, y_2, y_3\} \rightarrow$ Majority vote

+ + + $\rightarrow y_q$ = +ve

+ + - $\rightarrow y_q$ = -ve

K = odd (preferable)

$x_q \rightarrow$ k-NN to x_q

x_1 x_2 ... x_k

$\downarrow$

y_1 y_2 ... y_k

$\downarrow$

doing Majority vote

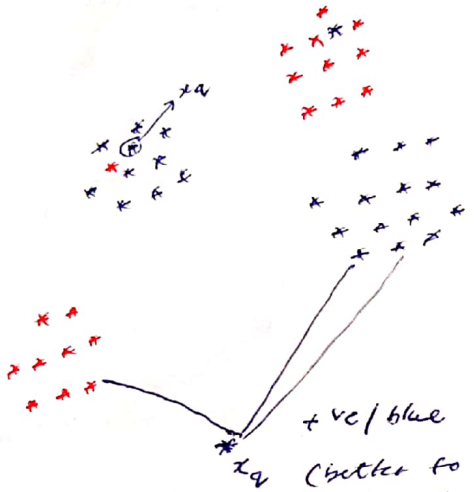
$\downarrow$

y_q

Simple idea of KNN

Failure cases of KNN -

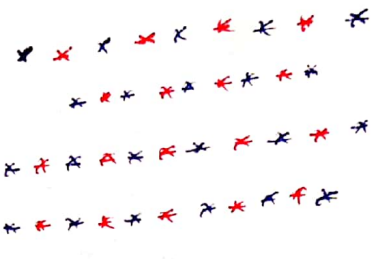
$k=3$



+ve / blue

x_q (better to say don't know)

far away from pts in D



$\downarrow$

jumble of +ve / -ve

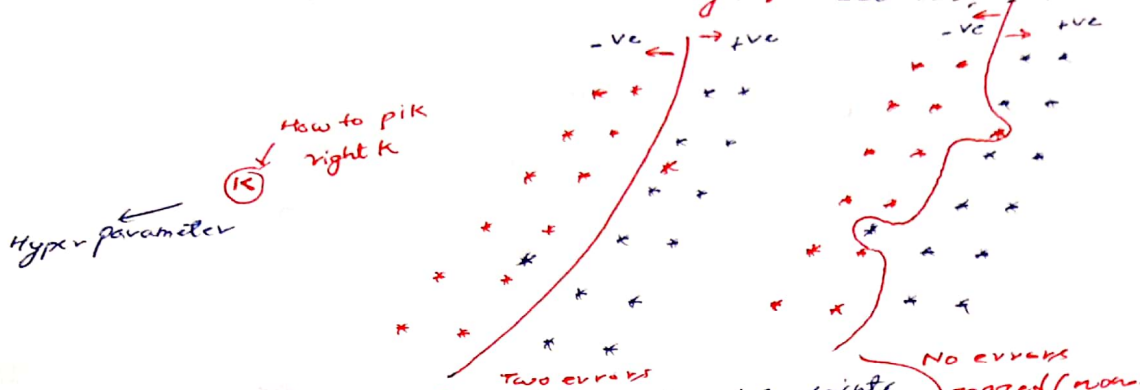
Here data is randomly spread

Here the KNN Fails.

Campus: PCE. Course: BTECH
Name of Faculty: Praveen Kumar Yadav
Class/Section: III CSE - A
Name of Subject: ML
Date: 26-02-24
Code: 6CS4-01

Decision surface for K-NN as K changes:-

Let K = 3, 5, 7, 15



Draw the line/curve to seperate the data points

Two errors

No errors

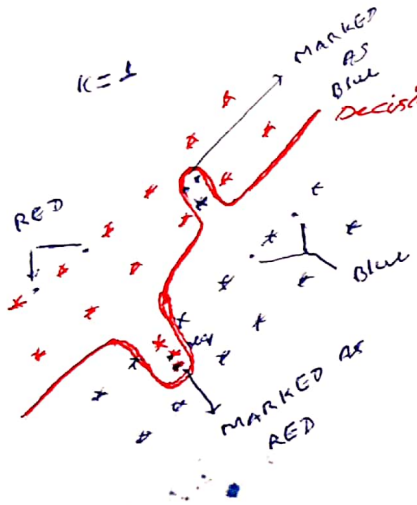
Jagged/non-smooth curve

Decision Surfaces or Decision Curves (Seperating region with curves)

Decision Surfaces or Decision Curve

Let K = 5

Let K = 1



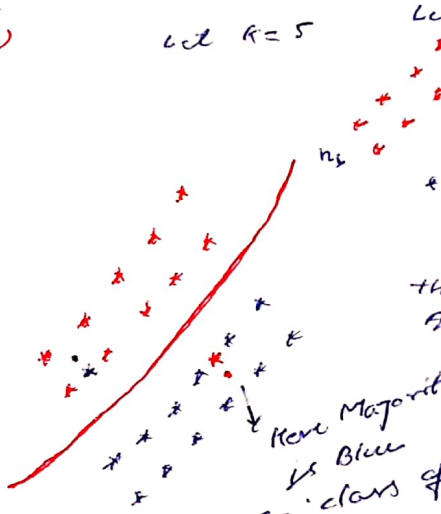
Decision Surfaces overfitting

Let K = L

Here, Jagged/non smooth curve

Let K = n

Total no of points.



If $n_1 > n_2$ that means Any xq becomes always red.

Here Majority is Blue so class of that point is Blue

n_1 = 1000
n_2 = 600
n_3 = 400
$n_1 > n_2$

Smoother Curve/surfaces

As K increases the Smoothness Curve of increases.

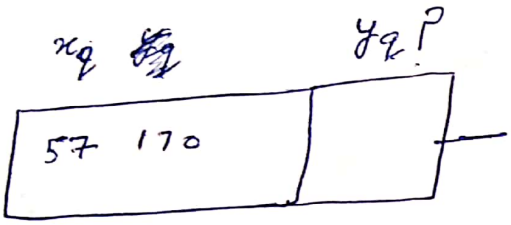
under fitting

everything becomes the majority class.

Numerical Example -

Weight	Height	CLASS	Euclidean distance d_{6-7}
51	167	UNDERWEIGHT	13
62	182	NORMAL	13.4
69	176	NORMAL	7.6
64	173	NORMAL	8.2
65	172	UNDERWEIGHT	4.1
56	174	NORMAL	1.4
58	169	NORMAL	3
57	173	NORMAL	2
55	170		

K=3



we have to classify the below set as normal. since Majority is NORMAL.

Campus: PCE. Course: BTECH
Name of Faculty: Praveen Kumar Yadav
Class/Section: VI CSE-A
Name of Subject: ML
Date: 24-02-21
Code: CSPC61

$D = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ i \\ \vdots \\ n \end{bmatrix} \begin{bmatrix} \leftarrow x_i \rightarrow \end{bmatrix} \begin{bmatrix} y_i \end{bmatrix}$

$y_i \in \{0, 1\}$ → -ve, +ve

$x_q \rightarrow y_q$

Given Dataset D_w → D_{Train} (70%), D_{Test} (30%) — do it randomly

such that $D_{Train} \cup D_{Test} = D_w$

$D_{Train} \cap D_{Test} = \phi$

$D_{Train} = n_1$ points

$D_{Test} = n_2$ points

$n_1 + n_2 = n$

$n_1 \approx 0.7\,n$

$n_2 \approx 0.3\,n$

$D_{Train} \rightarrow$ KNN → performance ?

D_w → n_1 → D_{Train} $(x_i, y_i)_{i=1}^{n_1}$

D_w → n_2 → D_{Test} $(x_i, y_i)_{i=1}^{n_2}$

Take each point x_i in $D_{Test} = \{(x_i, y_i)_{i=1}^{n_2}\}$

for each pt in D_{Test}

→ x_q = pt

→ use D_{Train} & KNN to predict y_q

---

for each point in D_{Test}:

x_q = pt

use D_{Train} & KNN to predict y_q from Majority vote

if $y_q == y_{pt}$

count += 1

End of FOR

count = # points for which D_{Train} KNN gave a correct ANSWER/CLASS LABEL

Accuracy = count/n_2 → no of points in my D_{Test}

Elbow Method