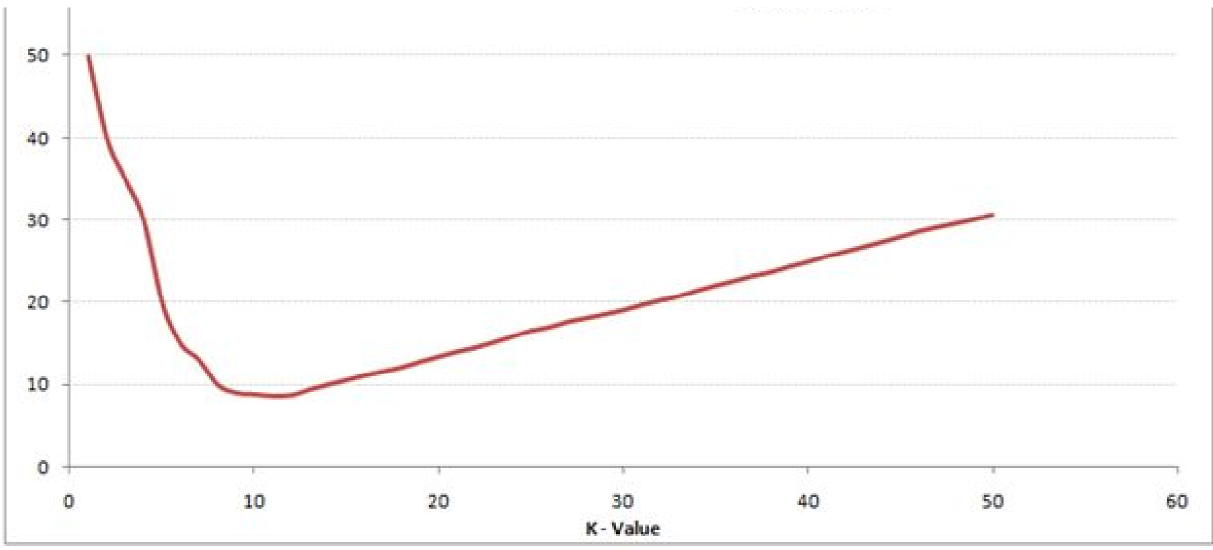
1.

a). Training error decreases when k moves from n to 1.

To be practical there is nothing to do much in the training phase of K-NN classification.
K=1 means the test point is declared as the same nearest data point always. So the error is zero. As K goes on increasing the other class comes into the picture which might dominate the true output class that brings error.

b).

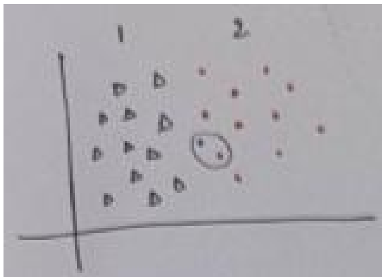


This is the graph of Generalization error vs k.
When k is small the output reflects easily with noisy points, this shows error is high.
When k is high output is biased to the class which has more number of data points irrespective of input.
The optimal value of k is in the mid range which actually predicts the output with less error.

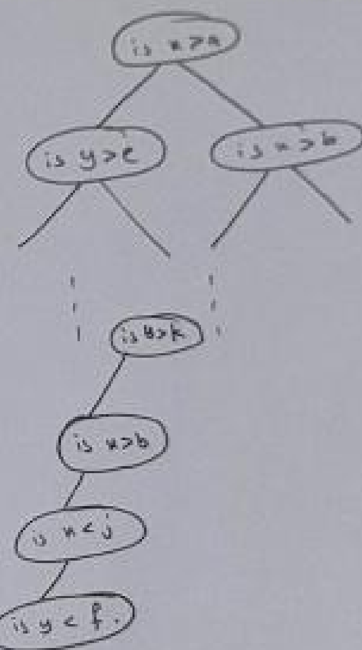
c). When the input dimension is high it gets very difficult for the algorithm to calculate the distance in all dimensions which degrades the performance of the algorithm.
A large amount of space is also required for high-dimensional data. The model has to store the whole data all the time.

d). Yes

In the 1-NN Classification Method, the euclidean distance is calculated between the new test data point and all the data points thereby classifying it to a particular class depends on the nearest point
In the Decision tree, we can reach the nearest point of the test data set by using several branches and nodes of the form (x>a, x>b, y<c,y<d).



The nearest point is decided by calculating Euclidean distances between test point and all the data points. Thereby classifying to class which has less distance i.e the nearest point.



This stage of tree gives us only one nearest point related with the test point and thereby classified as it belongs to class 2.

2.

a).

(2)

a) class 1 : $\{0.5, 0.1, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.35, 0.25\}$

class 2 : $\{0.9, 0.8, 0.75, 1.0\}$

$\sigma_1^2 : 0.0149, \quad \sigma_2^2 : 0.0092$

$$p(\text{class 1} \mid x : 0.6) = \frac{p(x = 0.6 \mid \text{class 1}) \cdot p(\text{class 1})}{p(x : 0.6)} \quad \text{— (1)}$$

$$p(x : 0.6 \mid \text{class 1}) : \frac{1}{\sqrt{2\pi}\,\sigma_1} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\mu_1 : \frac{2.6}{10} : 0.26 \qquad \mu_2 = \frac{3.45}{4} : 0.8625$

$$P_1 : \sum_{\substack{x = 0.5 \\ 0.1 \\ \vdots \\ 0.25}} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{\sigma_1^2}} : 0.714.$$

$$P_2 : \sum_{\substack{x = 0.9 \\ \vdots \\ 1.0}} \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(x-\mu_2)^2}{\sigma_2^2}} = 0.2857.$$

By using these values in (1)

$$P(\text{class 1} \mid x = 0.6) = 0.6305.$$

b).

2. b).

goal	football	golf	defence	offence	wic	offi	Strategy	Target (class)
1	0	1	1	1	0	1	1	P
0	0	0	1	0	0	1	1	P
1	0	0	1	1	0	1	0	P
0	1	0	0	1	1	0	1	P
0	0	0	1	1	0	1	1	P
0	0	0	1	1	0	0	1	P.
1	1	0	0	0	0	0	0	S
0	0	1	0	0	0	0	0	S
1	1	0	1	0	0	0	0	S.
1	1	0	1	0	0	0	1	S
1	1	0	1	1	0	0	0	S.
0	0	0	1	0	1	0	0	S.

Test data Point

Politics / . $(1,0,0,1,1,1,1,0)$ ~~Sport~~ $\rightarrow$ likelihood.

$$P\left(\text{Politics} \mid (1,0,0,1,1,1,1,0)\right) = \frac{\overbrace{P\left((1,0,0,1,1,1,1,0) \mid \text{Politics}\right)}^{\text{likelihood}} \times P(\text{Politics})}{P(1,0,0,1,1,1,1,0)}$$

Denominator is same, $P(\text{Politis}) = P(\text{sports})$.

So if we go with likelihood it is enough to give Solution.

goal	Polity	Sport
1	2/6	4/6
0	4/6	2/6

football	Polity	Sport
1	1/6	4/6
0	5/6	2/6

golf	Polity	Sport
1	1/6	1/6
0	5/6	5/6

defence	Polity	Sport
1	5/6	4/6
0	1/6	2/6

offence	Polity	Sport
1	5/6	1/6
0	1/6	5/6

wicket	Polity	Sport
1	1/6	1/6
0	5/6	5/6

office	Polity	Sport
1	4/6	0/6
0	2/6	6/6

strategy	Polity	Sport
1	5/6	1/6
0	1/6	5/6

1,0,0,1,1,1,1,0

$P(\text{goal}=1 / \text{Polity}) = 2/6$

$P(\text{football}=0 / \text{Polity}) = 5/6$

$P(\text{golf}=0 / \text{Polity}) = 5/6$

$P(\text{defence}=1 / \text{Polity}) = 5/6$

$P(\text{offence}=1 / \text{Polity}) = 5/6$

$P(\text{wicket}=1 / \text{Polity}) = 1/6$

$P(\text{office}=1 / \text{Polity}) = 4/6$

$P(\text{strategy}=0 / \text{Polity}) = 1/6$

$P(\text{Polity}) = 1/2$

$$P(\text{Polity} \mid (x=(1,0,0,1,1,1,1,0)) = P(x=(1,0,0,1,1,1,1,0) \mid \text{Polity}) \times P(\text{Polity})$$

$$= \frac{2}{6} \times \frac{5}{6} \times 5/6 \times 5/6 \times 5/6 \times 1/6 \times 4/6 \times 1/6 \times \frac{1}{2}$$

$$= 1.48 \times 10^{-3} = \underline{0.00148}$$

$$P(\text{sport} \mid x=(1,0,0,1,1,1,1,0)) \quad \frac{4}{6} \times \frac{2}{6} \times 5/6 \times \frac{4}{6} \times 1/6 \times 1/6 \times 0/6 \times 1/6$$

$$= \underline{0}$$

So the given data belongs to politics with probability (0.00148) [Neglecting denominator]