Assignment - 1

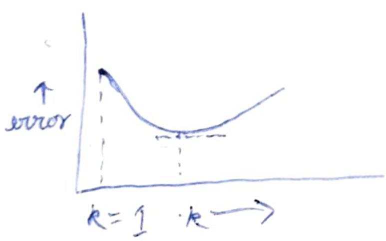
AI20BTECH11006

1)

a) $k = 1$

The training error will be 0, because the nearest point will be that instance itself.
As k begins to increase from 1 to n-1, we will start getting nearest points from other classes because the data is overlapping.
At $k = n$, we would get an error rate of 0.5 because the nearest n-points would consist of both classes of n/2 points each. even if we consider weighted

b) For $k = 1$ the generalisation error would be very high because of over-fitting. As k grows the error should decrease because in k ~~the~~ nearest neighbour the data point is assumed to be the same label as its neighbour. After a certain value of k, the error would start increasing again because more & more points of the other classes would be in the k-nearest neighbour.
The plot would look like



For the given question however the plot can fluctuate a lot because the points of other classes would be in the nearest neighbours even for small k.

c)

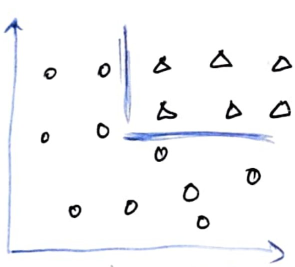
1) The ~~test~~ data size must increase a lot to accomodate the higher input dimension. This is because the distance can be similar & large for two different classes..

In certain dimensions, a set of points may be closer while overall points may be closer to each other. The points may live on a hypersphere.

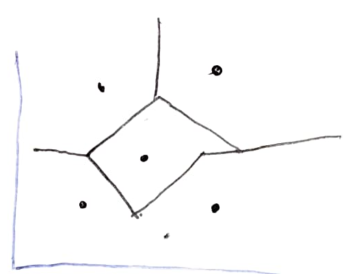
2) The amount of computation would increase linearly with the dimensions. ~~This~~

d) No,

The univariate decision tree can only form perpendicular lines to distinguish between classes, while in the case of 1-NN we have voronoi diagram which may not have only perpendicular lines. A pictorial representation for the same can be found below



univariate ~~1-NN~~ decision tree



1- NN

2)

a) The PDF of gaussian is ~~dif~~ given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

We have to maximise $f(x_1) f(x_2) \dots f(x_n)$

or $\log(f(x_1)) + \log(f(x_2)) + \dots + \log(f(x_n))$

The derivative wrt x should be 0.

$$\frac{f'(x_1)}{f(x_1)} + \frac{f'(x_2)}{f(x_2)} + \dots + \frac{f'(x_n)}{f(x_n)} = 0$$

$$x_1 - \mu + x_2 - \mu + x_3 - \mu + \dots + x_n - \mu = 0$$

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

for Class 1 $\Rightarrow$ $\mu_1 = \frac{0.5+0.1+0.2+0.4+0.3+0.2+0.1+0.35+0.25 \overset{+0.2}{}}{10}$

$= \cancel{\frac{2.6}{10}}$ ~~2.6~~ 0.26

for Class 2 $\Rightarrow$ $\mu_2 = \frac{0.9+0.8+0.75+1.0}{4} = \frac{3.45}{4} = 0.8625$

$\sigma_1^2 = 0.0149$

$\sigma_2^2 = 0.0092$

Class Probability

$$P_1 = \frac{n(\text{class } 1)}{\text{Total}} = \frac{5}{7}$$

$$P_2 = \frac{n(\text{class } 2)}{\text{Total}} = \frac{2}{7}$$

~~P(class 1~~

$$P(\text{class } 1 \mid X=0.6) = \frac{P(X=0.6 \mid \text{class } 1)\ P(\text{class } 1)}{P(X=0.6)}$$

$$= \frac{P(X=0.6 \mid \text{class } 1)\, p_1}{P(X=0.6, \text{class } 1) + P(X=0.6, \text{class } 2)}$$

$$= \frac{P(x=0.6 \mid \text{class } 1)\, p_1}{P(x=0.6 \mid \text{class } 1)\, p_1 + P(x=0.6 \mid \text{class } 2)\, p_2}$$

$$= \frac{\frac{1}{\sigma_1} e^{-\frac{(0.6-0.26)^2}{2\sigma_1^2}} \frac{5}{7}}{\frac{1}{\sigma_1} e^{-\frac{(0.6-0.26)^2}{2\sigma_1^2}} \frac{5}{7} + \frac{1}{\sigma_2} e^{-\frac{(0.6-0.8625)^2}{2\sigma_2^2}} \frac{2}{7}}$$

$$= \frac{\frac{1}{\sqrt{0.0149}} e^{-\frac{0.34^2}{2(0.0149)}} \times 5}{\frac{1}{\sqrt{0.0149}} e^{-\frac{0.34^2}{2(0.0149)}} 5 + \frac{1}{\sqrt{0.0092}} e^{-\frac{(0.2625)^2}{2(0.0092)}} 2}$$

$$\boxed{= 0.6319}$$

6)

$$x_{politics} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$x_{sports} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$x = [1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0]$$

The required probability is

$$\frac{\frac{1}{3} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{4}{6} \times \frac{1}{6}}{\frac{1}{3} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{4}{6} \times \frac{1}{6} + \frac{4}{6} \times \frac{2}{6} \times \frac{5}{6} \times \frac{4}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{0}{6} \times \frac{5}{6}}$$

$$\boxed{= 1}$$