2A) a) Given training data from class '1' are

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}

mean of class 1 = $\frac{\Sigma x_i}{N_{c_1}}$ $x_i \in$ class 1

$N_{c_1} \rightarrow$ no. of elements in class 1

$\mu_1 = \frac{2.6}{10} = 0.26$

mean of class 2 = $\frac{\Sigma y_i}{N_{c_2}} = \frac{3.45}{4} = 0.8625 = \mu_2$

variance in class 1 = $\sigma_1 = 0.0169$

variance in class 2 = $\sigma_2 = 0.0092$

Assumin gaussian distribution over the data in both class the pdf of both classes are given by the expressions

$$P(D/C_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{x_i - \mu_1}{2\sigma_1^2}\right) \rightarrow \text{likelihood of } C_1$$

$$P(D/C_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{x_i - \mu_2}{2\sigma_2^2}\right) \rightarrow \text{likelihood of Data given class } C_2$$

Probability Distribution over the two classes can be assumed to be following Bernoulli distribution

$$\text{Pdf} = p^{n_1}(1-p)^{n_2}$$

where $n_1 \rightarrow$ no. of instances of class 1

$n_2 \rightarrow$ no. of instances in

The maximum likelihood estimate of the ~~pro~~ parameter is given by maximizing the likelihood expression.

the estimate is $p = \frac{n_1}{n_1+n_2}$, $(1-p) = \frac{n_2}{n_1+n_2}$

$P(C_1) = p = \frac{n_1}{n_1+n_2} = \frac{10}{10+4} = \frac{10}{14} = \cancel{0.28}\ 0.7143$

$P(C_2) = \frac{4}{14} = 0.2857$

$\cancel{P\left(x=0.6/C_1\right) = P(C_1}$

$$\boxed{P(C/x) = \frac{P(x/C).P(C)}{\sum_i P(x/C_i).P(C_i)}}$$

(i) $P(C_1/x=0.6) \propto P(x=0.6/C_1).P(C_1)$

$$P(x=0.6/C_1) = \frac{1}{\sqrt{2\pi(0.0149)^2}} \exp\left(\frac{(0.6-0.26)^2}{2(0.0149)^2}\right)$$

$= 0.06756$

$P(C_1/x=0.6) \propto (0.06756)(0.7143) = \cancel{0.6305}\ 0.0425811$

$P.$

$$P(C_1/x=0.6) = \frac{0.0425811}{\sum_i P(x/C_i)P(C_i)} = 0.6305.$$

Scanned with CamScanner

(b) Given the two documents

$x = $ (goal, football, golf, defence, offence, wicket, office, strategy)

P $C_1 = $ Politics, $C_2 = $ Sport.

$P(C_1) = 1/2$; $P(C_2) = 1/2$ ($\because$ the classes are only two, we can assume Bernoulli prior on both the classes).

and max likelihood estimate over the classes $= p = 6/12 = 1/2$

$1-p = 1/2$

$D = (1,0,0,1,1,1,1,0) =$

Since one of the word's frequency is '0' in document 'politics' we apply smoothening

$$P(C_1/D) = \frac{P(D/C_1) \cdot P(C_1)}{\sum_i P(C_i) \cdot P(D/C_i)}$$

, $P(C_1) = 1/2$

$P(C_2) = 1/2$

$$P(D/C_1) = \left(\frac{3}{8}\right)\left(\frac{2}{8}\right)\left(\frac{2}{8}\right)\left(\frac{6}{8}\right)\left(\frac{2}{8}\right)\left(\frac{5}{8}\right)\left(\frac{6}{8}\right)\left(\frac{6}{8}\right)$$

$$P(D/C_2) = \left(\frac{5}{8}\right)\left(\frac{5}{8}\right)\left(\frac{2}{8}\right)\left(\frac{5}{8}\right)\left(\frac{2}{8}\right)\left(\frac{2}{8}\right)\left(\frac{1}{8}\right)\left(\frac{2}{8}\right)$$

$$P(C_1/D) = \frac{P(D/C_1) \cdot P(C_1)}{P(D/C_1) \cdot P(C_1) + P(D/C_2) \cdot P(C_2)}$$

$= 0.878.$