Homework - 1

Q2)

a) Here we have two classes c_1 and c_2.

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}

→ Using Bayes theorm

$$P(c_j/x) = \frac{P(x/c_j)\,P(c_j)}{P(x)}$$

$$\approx \frac{P(x/c_j)\,P(c_j)}{\sum_{k=1}^{K} P(x/c_k)\,P(c_k)}$$

$$P(x/c_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}$$

Given → $\sigma_1^2 = 0.0149$

$\sigma_2^2 = 0.0092$

$\mu_1 = (0.5 + 0.1 + 0.2 + 0.4 + 0.3 + 0.2 + 0.2 + 0.1 + 0.35 + 0.25)/10$

$\mu_1 = 0.026$

$\mu_2 = (0.9 + 0.8 + 0.75 + 1)/4$

$\mu_2 = 0.8625$

Now, class prob

$P(c_1) = \frac{10}{10+4}$

$P(c_1) = 0.714$

$P(c_2) = \frac{4}{14}$

$P(c_2) = 0.2857$

$$P(x|c_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\frac{(0.6-0.26)^2}{0.0149}}$$

$$= \frac{1}{0.305972} * 0.020667$$

$$\boxed{P(x|c_1) = 0.067549}$$

$$P(x|c_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{(0.6-0.8625)^2}{0.0092}}$$

$$= \frac{1}{0.24042} * 0.02363$$

$$\boxed{P\left(\frac{x}{c_2}\right) = 0.0982863}$$

Now, probability of having 0.6 in class c_1

$$P(c_1|0.6) = \frac{0.0675 * 0.714}{0.714 * 0.0675 + 0.098286 * 0.2857}$$

$$\boxed{P(c_1|0.6) = 0.631855}$$

$$\boxed{P(c_2|0.6) = 0.3681}$$

So, 0.6 point belong to class 1.

→ Class probability P1 and P2 are

$$\boxed{P(c_1) = \frac{10}{14}} = 0.714$$

$$\boxed{P(c_2) = \frac{4}{14}} = 0.2857$$

Q2)

b) The formulae for Naive Bayes is

$$P(c_j|x) = \frac{P(x|c_j)\,P(c_j)}{\sum_{k=1}^{K} P(x|c_k)\,P(c_k)}$$

→ where k is number of classes

Given a document $x = (1,0,0,1,1,1,1,0)$

→ where 1 represent the presence of a word and 0 means absence of that word.

→ They are using 8 attribute for classification

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

$$\text{Prob of class} = \frac{\text{Number of doc in that class}}{\text{Total Number of doc}}$$

$$\text{Prob (Politics)} = \frac{1}{2}, \quad \text{Prob (sport)} = \frac{1}{2}$$

→ In Naive bayes, prob of all words are independent

$$\text{Prob}(x_i|\text{class}) = \frac{\text{Number of occurence of } x_i \text{ in that class}}{\text{Total words in that class}}$$

P.T.O

$P(\text{goal} \mid \text{politics}) = \frac{2}{24}$

$P(\text{defence} \mid \text{politics}) = \frac{5}{24}$

$P(\text{offence} \mid \text{politics}) = \frac{5}{24}$

$P(\text{wicket} \mid \text{politics}) = \frac{1}{24}$

$P(\text{office} \mid \text{politics}) = \frac{4}{24}$

Thus, $\boxed{Prob(x \mid Politics) = \frac{2 * 5 * 5 * 1 * 4}{(24)^5}}$

$P(\text{goal} \mid \text{sports}) = 4/16$

~~$P(\text{defence} \mid \text{politics}) = 4/16$~~

$P(\text{defence} \mid \text{sports}) = 4/16$

$P(\text{offence} \mid \text{sports}) = 1/16$

$P(\text{wicket} \mid \text{sports}) = 1/16$

$P(\text{office} \mid \text{sports}) = 0/16$

$$\boxed{Prob(x \mid sports) = 0}$$

So,

$$Prob(Politics \mid x) = \frac{2 * 5 * 5 * 1 * 4 * 24^5}{(24)^5 * 2 \times 5 \times 5 * 1 \times 4}$$

$$\boxed{Prob(Politics \mid x) = 1}$$