CSCI 5521 - INTRO TO MACHINE LEARNING

TANVEER SINGH VIRDI

ASSIGNMENT 1

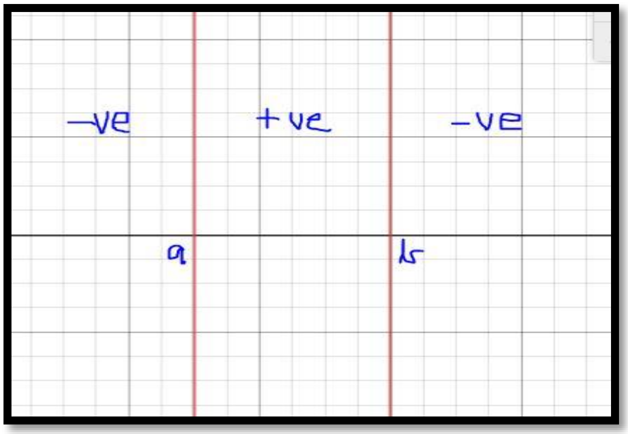
SEPTEMBER 29, 2019

PROBLEM 1:

SOLUTION:

Given,

Target function specified by an interval [a,b] in $\mathbb{R}$. Any example lying within the target function is labeled positive. Otherwise it is labelled negative. This can be visualized as :



VC Dimension is defined as the maximum number of points that can be shattered by a hypothesis $\mathcal{H}$.
The VC Dimension, VC($\mathcal{H}$) of the given target function is 2. This can be explained as follows :
Lets take any 3 numbers on the real number line $\mathbb{R}$, (p, q, r) such that $p < q < r$ and $p < a$; $a < q < b$ and $r > b$. Number q will be labeled as positive since it lies between a and b on the number line. Also, p and r would be labeled negative. We can see that any such 3 numbers the target function is not shattered.
For 2 numbers on the real number line $\mathbb{R}$, (p, q) , these numbers can be taken in three possible ways : *either* i)both of these numbers lie within the interval [a,b] ,ii) one of the two lies in the interval [a,b] and the other does not *or* iii) none of the two numbers lie within the interval [a,b] . In this case we see that the target function is shattered because we can successfully classify all possible combinations of the two numbers with the given interval.
Thus the VC dimension of the target function is 2.

PROBLEM 2:

SOLUTION:

a. Maximum likelihood estimation(MLE) for the function :

$$f(\mathcal{X}|\theta) = \frac{1}{\theta-1} e^{-\frac{x}{\theta-1}} \text{ , x > 0, } \theta > 1$$

Log likelihood of the given function is :

$$\mathcal{L}(\theta|x) = \log l(\theta|x) = \sum_{t=1}^{N} \log f(x^t|\theta)$$

$$\Rightarrow \; \sum_{t=1}^{N} \log\left(\frac{1}{\theta-1} e^{-\frac{x^t}{\theta-1}}\right)$$

$$\Rightarrow \; \sum_{t=1}^{N} [\log\left(\frac{1}{\theta-1}\right) + \log\left(e^{-\frac{x^t}{\theta-1}}\right)]$$

$$\Rightarrow \; \sum_{t=1}^{N}[\log\left(\frac{1}{\theta-1}\right) - \frac{x^t}{\theta-1}]$$

$$\Rightarrow \; \sum_{t=1}^{N}[-\log(\theta-1) - \frac{x^t}{\theta-1}]$$

$$\Longrightarrow \; -N \log(\theta-1) - \frac{1}{\theta-1} \sum_{t=1}^{N} x^t$$

θ that maximizes the log likelihood can be found by solving for $\frac{d\mathcal{L}}{d\theta} = 0$:

$$\frac{d\mathcal{L}}{d\theta} = -\frac{N}{\theta-1} + \frac{1}{(\theta-1)^2} \sum_{t=1}^{N} x^t = 0$$

$$\Longrightarrow \; N(\theta-1) = \sum_{t=1}^{N} x^t$$

$$\Longrightarrow \; \theta = 1 + \frac{\sum_{t=1}^{N} x^t}{N}$$

Hence the maximum likelihood estimate is :

$$\boxed{\hat{\theta} = 1 + \frac{\sum_{t=1}^{N} x^t}{N}}$$

b. Maximum likelihood estimation(MLE) for the function :

$$f(\mathcal{X}|\theta) = (\theta-1)x^{(\theta-2)}, 0 \leq \mathrm{x} \leq 1, 1 < \theta < \infty$$

Log likelihood of the given function is :

$$\mathcal{L}(\theta|x) = \log l(\theta|x) = \sum_{t=1}^{N} \log f(x^t|\theta)$$

$$\Rightarrow \; \sum_{t=1}^{N} \log((\theta-1)x^{(\theta-2)})$$

$$\Rightarrow \; \sum_{t=1}^{N}[\log(\theta-1) + (\theta-2)\log x^t]$$

$$\Rightarrow \; N\log(\theta-1) + (\theta-2)\sum_{t=1}^{N} \log x^t]$$

θ that maximizes the log likelihood can be found by solving for $\frac{d\mathcal{L}}{d\theta} = 0$:

$$\frac{d\mathcal{L}}{d\theta} = \frac{N}{(\theta-1)} + \sum_{t=1}^{N} \log x^t = 0$$

$$\Rightarrow \frac{\sum_{t=1}^{N} \log x^t}{N} = \frac{1}{(1-\theta)}$$

$$\Rightarrow \theta = 1 - {1}\Big/{\frac{\sum_{t=1}^{N} \log x^t}{N}}$$

Hence the maximum likelihood estimate is :

$$\boxed{\hat{\theta} = 1 - ({1}\Big/{\frac{\sum_{t=1}^{N} \log x^t}{N}})}$$

c. Maximum likelihood estimation(MLE) for the function :

$$f(\mathcal{X}|\theta) = \frac{1}{\theta} \quad ,0 \le \mathrm{x} \le \theta$$

Likelihood of the parameter θ is :

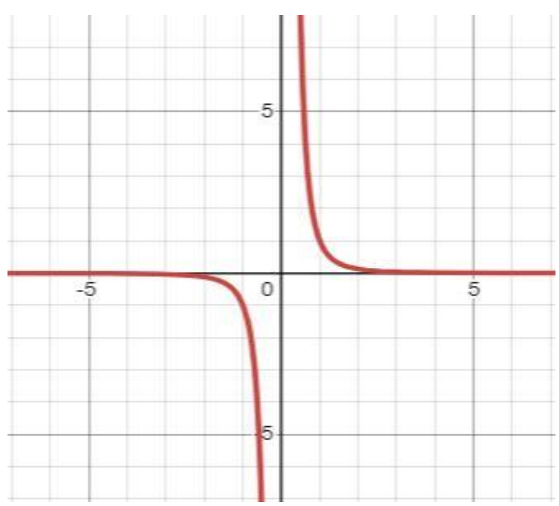
$$l(\theta|x) = \prod_{t=1}^{N} f(x^t|\theta)$$

$$\Rightarrow \prod_{t=1}^{N} \frac{1}{\theta} = \frac{1}{\theta^{\mathrm{N}}}$$

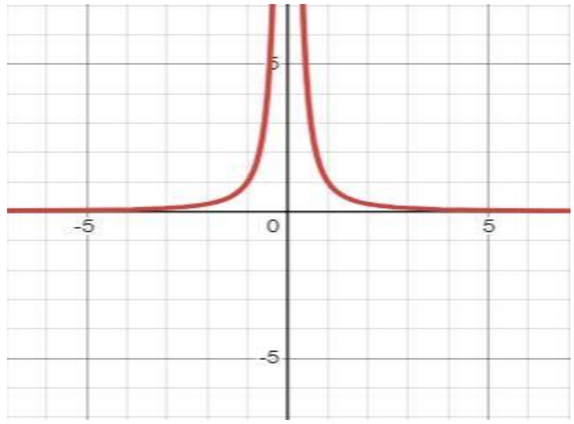
θ that maximizes the log likelihood can be found by solving for $\frac{dl}{d\theta} = 0$:

$\frac{dl}{d\theta} = -\frac{\mathrm{N}}{\theta^{\mathrm{N+1}}} = 0$ (We can't find a definite value of θ with this)

Drawing a graph for likelihood function we see that for odd values of N, (N+1) is even and the graph of the function $\frac{1}{\theta^{\mathrm{N}}}$ looks like :



For even values of N, (N+1) is odd and the graph of the function $\frac{1}{\theta^{N}}$ looks like :



To maximize the given function $\frac{1}{\theta^{N}}$, we need the smallest value of θ. We know that $0 \leq x \leq \theta.$

This implies that $\theta \geq \ x_t$ for all t= 1 to N. Thus we need the smallest value of θ satisfying $\theta \geq \ x_t$.

$$\boxed{\hat{\theta} \ = \ \max\,(x_t) \text{ for all t= 1 to N}}$$

PROBLEM 3:

SOLUTION:

a. Given the priors P(C1) and P(C2) for the classes C1 and C2.

Also,

Bernoulli densities: p1 = P(x = 0 | C1) , p2 = P(x = 0 | C2)

Then, Bernoulli densities P(x = 1 | C1) = 1-p1 and P(x = 1 | C2) = 1-p2

The probability of classifying x = 0 as C1 :

$$P(\ C1 \mid x = 0) \ = \frac{P(C1)\,P(\ x{=}0\ |C1)}{P(x{=}0)} = \frac{P(C1)\,P(\ x{=}0\ |C1)}{P(x{=}0\ |C1)\,P(C1) + P(x{=}0\ |C2)\,P(C2)}$$

$$= \frac{P(C1) * p1}{p1 * P(C1) + p2 * P(C2)}$$

The probability of classifying x = 0 as C2 :

$$P(C2 \mid x=0) = \frac{P(C2)\,P(x=0\,|C2)}{P(x=0)} = \frac{P(C2)\,P(x=0\,|C2)}{P(x=0\,|C1)\,P(C1)+P(x=0\,|C2)\,P(C2)}$$

$$= \frac{P(C2)*p2}{p1*P(C1)+p2*P(C2)}$$

Then we will classify x=0 as Class C1 if :

$$P(C1 \mid x=0) > P(C2 \mid x=0) \implies \frac{P(C1 \mid x=0)}{P(C2 \mid x=0)} > 1$$

Otherwise we will classify x=0 as Class C2.

Similarly,

The probability of classifying x = 1 as C1 :

$$P(C1 \mid x=1) = \frac{P(C1)\,P(x=1\,|C1)}{P(x=1)} = \frac{P(C1)\,P(x=1\,|C1)}{P(x=1\,|C1)\,P(C1)+P(x=1\,|C2)\,P(C2)}$$

$$= \frac{P(C1)*(1-p1)}{(1-p1)*P(C1)+(1-p2)*P(C2)}$$

The probability of classifying x = 1 as C2 :

$$P(C2 \mid x=1) = \frac{P(C2)\,P(x=1\,|C2)}{P(x=1)} = \frac{P(C2)\,P(x=1\,|C2)}{P(x=1\,|C1)\,P(C1)+P(x=1\,|C2)\,P(C2)}$$

$$= \frac{P(C2)*(1-p2)}{(1-p1)*P(C1)+(1-p2)*P(C2)}$$

Then we will classify x=1 as Class C1 if :

$$P(C1 \mid x=1) > P(C2 \mid x=1) \implies \frac{P(C1 \mid x=1)}{P(C2 \mid x=1)} > 1$$

Otherwise we will classify x=1 as Class C2.

b. Given,
D-dimensional Bernoulli densities specified by $p_{ij} \equiv p(x_j = 0|C_1)$, i = 1,2 and j = 1,2,....,D.

For D dimensional Bernoulli densities specified by $p_{ij} \equiv p(x_j = 1|C1)$, the probability distribution $P(x|C_i)$ is given by :

$$P(x|C1) = \prod_{j}^{D} {p_{ij}}^{x_j}(1-p_{ij})^{1-x_j}$$

Conversely, for D dimensional Bernoulli densities specified by $p_{ij} \equiv p(x_j = 0|C_1)$, the probability distribution $P(x|C_i)$ will given by :

$$\boxed{P(x|C1) = \prod_{j}^{D}(1-p_{ij})^{x_j} {p_{ij}}^{1-x_j}}$$

So, for class C1 :

$$P(x|C1) = \prod_{j}^{D}(1-p_{1j})^{x_j} {p_{1j}}^{1-x_j}$$

And for class C2 :

$$P(x|C2) = \prod_{j}^{D}(1-p_{2j})^{x_j} {p_{2j}}^{1-x_j}$$

The posterior probabilities for these two classes will be :

$$P(C1|x) = \frac{P(C1)P(x|C1)}{P(x)} = \frac{P(C1)\ P(\ x\ |C1)}{P(x\ |C1)\ P(C1) +\ P(x\ |C2)\ P(C2)}$$

$$P(C2|x) = \frac{P(C2)P(x|C2)}{P(x)} = \frac{P(C2)\ P(\ x\ |C2)}{P(x\ |C1)\ P(C1) +\ P(x\ |C2)\ P(C2)}$$

To classify a sample into one of the classes C1 or C2, we compare the posterior probabilities.
We will classify a sample into class C1 if :

$$P(C1|x) > P(C2|x) \implies \frac{P(C1|x)}{P(C2|x)} > 1 \implies \frac{P(C1)\ P(x\ |C1)}{P(C2)P(x\ |C2)} > 1 \implies \frac{P(C1)\ \prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j}}{P(C2)\ \prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j}}$$

Otherwise, we will classify it into class C2.
The classification rule for classifying a sample into a class C1 or C2 can be written as :

$$\boxed{\text{Class } C_i \quad if, \quad i = \underset{i=1,2}{\operatorname{argmax}} P(Ci) \prod_j^D (1-p_{ij})^{x_j} {p_{ij}}^{1-x_j}}$$

c. Given D =2, p_{11} = 0.6 , p_{12} = 0.1, p_{21} = 0.6, p_{22} = 0.9,

Also, given three different priors for the Classes C1 and C2:

P(C1) = 0.2 , 0.6, 0.8

P(C2) = 1 - P(C1) = 0.8, 0.4, 0.2

$$\boxed{P(\ x \mid C_i\ ) = \prod_j^D (1-p_{ij})^{x_j} {p_{ij}}^{1-x_j}}$$

For the class C1 :

P(x | C1) = $\prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j}$

Similarly for class C2 :

P(x | C2) = $\prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j}$

- **<u>For Sample (x1,x2) = (0,0):</u>**

P(x | C1) = $\prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j} = (1-0.6)^0(0.6)^1 * (1-0.1)^0(0.1)^1$

= 0.6 * 0.1 = 0.06

P(x | C2) = $\prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j} = (1-0.6)^0(0.6)^1 * (1-0.9)^0(0.9)^1$

= 0.6 * 0.9 = 0.54

- **P(C1) = 0.2 and P(C2) = 0.8**

P(C1 | x) = $\frac{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)}}{P(x)} = \frac{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)}}{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)} + \mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |\ C2)}}$

P(C1 | x) = $\frac{0.2*0.06}{(0.2*0.06+0.8*0.54)}$ = **0.027**

P(C2 | x) = $\frac{\mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |C2)}}{P(x)} = \frac{\mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |C2)}}{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)} + \mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |\ C2)}}$

P(C2 | x) = $\frac{0.8*0.54}{(0.2*0.06+0.8*0.54)}$ = **0.973**

- **P(C1) = 0.6 and P(C2) = 0.4**

P(C1 | x) = $\frac{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)}}{P(x)} = \frac{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)}}{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)} + \mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |\ C2)}}$

P(C1 | x) = $\frac{0.6*0.06}{(0.6*0.06+0.4*0.54)}$ = **0.143**

P(C2 | x) = $\frac{\mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |C2)}}{P(x)} = \frac{\mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |C2)}}{\mathrm{P(\ C1\ )} * \mathrm{P(\ x\ |\ C1)} + \mathrm{P(\ C2\ )} * \mathrm{P(\ x\ |\ C2)}}$

P(C2 | x) = $\frac{0.4*0.54}{(0.6*0.06+0.4*0.54)}$ = **0.857**

- **P(C1) = 0.8 and P(C2) = 0.2**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.8*0.06}{(0.8*0.06+0.2*0.54)} = \mathbf{0.307}$$

$$P(C2 \mid x) = \frac{P(C2) * P(x \mid C2)}{P(x)} = \frac{P(C2) * P(x \mid C2)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.2*0.54}{(0.8*0.06+0.2*0.54)} = \mathbf{0.693}$$

❖ **<u>For Sample (x1,x2) = (0,1):</u>**

$$P(x \mid C1) = \prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j} = (1-0.6)^0(0.6)^1 * (1-0.1)^1(0.1)^0$$

$$= 0.6 * 0.9 = 0.54$$

$$P(x \mid C2) = \prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j} = (1-0.6)^0(0.6)^1 * (1-0.9)^1(0.9)^0$$

$$= 0.6 * 0.1 = 0.06$$

- **P(C1) = 0.2 and P(C2) = 0.8**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.2*0.54}{(0.2*0.54+0.8*0.06)} = \mathbf{0.693}$$

$$P(C2 \mid x) = \frac{P(C2) * P(x \mid C2)}{P(x)} = \frac{P(C2) * P(x \mid C2)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.8*0.06}{(0.2*0.54+0.8*0.06)} = \mathbf{0.307}$$

- **P(C1) = 0.6 and P(C2) = 0.4**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.6*0.54}{(0.6*0.54+0.4*0.06)} = \mathbf{0.931}$$

$$P(C2 \mid x) = \frac{P(C2)*P(x \mid C2)}{P(x)} = \frac{P(C2)*P(x \mid C2)}{P(C1)*P(x \mid C1)+P(C2)*P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.4*0.06}{(0.6*0.54+0.4*0.06)} = \mathbf{0.069}$$

- **P(C1) = 0.8 and P(C2) = 0.2**

$$P(C1 \mid x) = \frac{P(C1)*P(x \mid C1)}{P(x)} = \frac{P(C1)*P(x \mid C1)}{P(C1)*P(x \mid C1)+P(C2)*P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.8*0.54}{(0.8*0.54+0.2*0.06)} = \mathbf{0.973}$$

$$P(C2 \mid x) = \frac{P(C2)*P(x \mid C2)}{P(x)} = \frac{P(C2)*P(x \mid C2)}{P(C1)*P(x \mid C1)+P(C2)*P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.2*0.06}{(0.8*0.54+0.2*0.06)} = \mathbf{0.027}$$

❖ **<u>For Sample (x1,x2) = (1,0):</u>**

$$P(x \mid C1) = \prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j} = (1-0.6)^1(0.6)^0 * (1-0.1)^0(0.1)^1$$

$$= 0.4 * 0.1 = 0.04$$

$$P(x \mid C2) = \prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j} = (1-0.6)^1(0.6)^0 * (1-0.9)^0(0.9)^1$$

$$= 0.4 * 0.9 = 0.36$$

- **P(C1) = 0.2 and P(C2) = 0.8**

$$P(C1 \mid x) = \frac{P(C1)*P(x \mid C1)}{P(x)} = \frac{P(C1)*P(x \mid C1)}{P(C1)*P(x \mid C1)+P(C2)*P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.2*0.04}{(0.2*0.04+0.8*0.36)} = \mathbf{0.027}$$

$$P(C2 \mid x) = \frac{P(C2)*P(x \mid C2)}{P(x)} = \frac{P(C2)*P(x \mid C2)}{P(C1)*P(x \mid C1)+P(C2)*P(x \mid C2)}$$

P(C2 | x) = $\frac{0.8*0.36}{(0.2*0.04+0.8*0.36)}$ = **0.973**

- **P(C1) = 0.6 and P(C2) = 0.4**

P(C1 | x) = $\frac{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)}}{P(x)} = \frac{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)}}{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)} + \mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}$

P(C1 | x) = $\frac{0.6*0.04}{(0.6*0.04+0.4*0.36)}$ = **0.143**

P(C2 | x) = $\frac{\mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}{P(x)} = \frac{\mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)} + \mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}$

P(C2 | x) = $\frac{0.4*0.36}{(0.6*0.04+0.4*0.36)}$ = **0.857**

- **P(C1) = 0.8 and P(C2) = 0.2**

P(C1 | x) = $\frac{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)}}{P(x)} = \frac{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)}}{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)} + \mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}$

P(C1 | x) = $\frac{0.8*0.04}{(0.8*0.04+0.2*0.36)}$ = **0.307**

P(C2 | x) = $\frac{\mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}{P(x)} = \frac{\mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}{\mathrm{P(\,C1\,)} * \mathrm{P(\,x \mid C1)} + \mathrm{P(\,C2\,)} * \mathrm{P(\,x \mid C2)}}$

P(C2 | x) = $\frac{0.2*0.36}{(0.8*0.04+0.2*0.36)}$ = **0.693**

❖ **<u>For Sample (x1,x2) = (1,1):</u>**

P(x | C1) = $\prod_j^D (1-p_{1j})^{x_j} {p_{1j}}^{1-x_j}$ = $(1\text{-}0.6)^1(0.6)^0 * (1\text{-}0.1)^1(0.1)^0$

= 0.4 * 0.9 = 0.36

P(x | C2) = $\prod_j^D (1-p_{2j})^{x_j} {p_{2j}}^{1-x_j}$ = $(1\text{-}0.6)^1(0.6)^0 * (1\text{-}0.9)^1(0.9)^0$

= 0.4 * 0.1 = 0.04

- **P(C1) = 0.2 and P(C2) = 0.8**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.2*0.36}{(0.2*0.36+0.8*0.04)} = \mathbf{0.693}$$

$$P(C2 \mid x) = \frac{P(C2) * P(x \mid C2)}{P(x)} = \frac{P(C2) * P(x \mid C2)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.8*0.04}{(0.2*0.36+0.8*0.04)} = \mathbf{0.307}$$

- **P(C1) = 0.6 and P(C2) = 0.4**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.6*0.36}{(0.6*0.36+0.4*0.04)} = \mathbf{0.931}$$

$$P(C2 \mid x) = \frac{P(C2) * P(x \mid C2)}{P(x)} = \frac{P(C2) * P(x \mid C2)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.4*0.04}{(0.6*0.36+0.4*0.04)} = \mathbf{0.069}$$

- **P(C1) = 0.8 and P(C2) = 0.2**

$$P(C1 \mid x) = \frac{P(C1) * P(x \mid C1)}{P(x)} = \frac{P(C1) * P(x \mid C1)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C1 \mid x) = \frac{0.8*0.36}{(0.8*0.36+0.2*0.04)} = \mathbf{0.973}$$

$$P(C2 \mid x) = \frac{P(C2) * P(x \mid C2)}{P(x)} = \frac{P(C2) * P(x \mid C2)}{P(C1) * P(x \mid C1) + P(C2) * P(x \mid C2)}$$

$$P(C2 \mid x) = \frac{0.2*0.04}{(0.8*0.36+0.2*0.04)} = \mathbf{0.027}$$

PROBLEM 4:

SOLUTION:

TABLE OF ERROR RATES ON THE VALIDATION SET

SIGMA VALUE	ERROR RATE
0.00001	0.54
0.0001	0.54
0.001	0.54
0.01	0.54
0.1	0.51
1	0.515
2	0.455
3	0.46
4	0.46
5	0.46
6	0.46

Error rate observed on the test dataset for sigma = 2 was 0.445.