CSCI-5521 - Homework-1

Name:- Sanjay Yellambalase Ravikumar
Id : yella016@umn.edu, 5397519

Q1) VC dimension of intervals in $\mathbb{R}$ = d_x.

Target function specified by an interval in [a, b].

i) One point can easily be shattered by target function whether it is +ve or -ve.

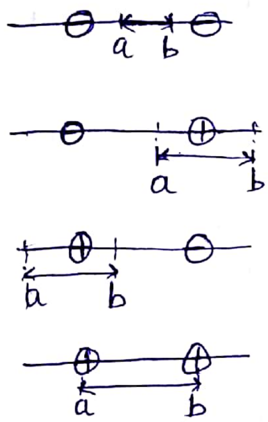
ii) For two points, we have 4 cases:

a)

b)

c)

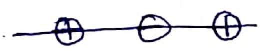
d)



⊖ - negative
⊕ - positive

Points inside [a, b] interval are positive and points outside are negative.

(iii) For three points, we cannot represent alternating points (+, -, +) due to convexity.



Cannot shatter 3 points using [a, b] interval in this case.

∴ Hence, VC dimension of intervals in $\mathbb{R}$ is 2.

Q2) Maximum Likelihood Estimation (MLE) for the pdfs.

(a) $f(x|\theta) = \frac{1}{\theta-1} e^{\frac{-x}{\theta-1}}$ $\quad x > 0, \theta > 1$

$$L_n(\theta) = \prod_{t=1}^{n} \frac{1}{\theta-1} \cdot e^{\frac{-x^t}{\theta-1}}$$

$$\log(L_n(\theta)) = \sum_{t=1}^{n} \log\left(\frac{1}{\theta-1} \cdot e^{\frac{-x^t}{\theta-1}}\right)$$

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

$$\frac{\partial(\log L_n(\theta))}{\partial\theta} = 0$$

$$\Rightarrow \sum_{t=1}^{n}\left[-\frac{1}{\theta-1} + \frac{x^t}{(\theta-1)^2}\right] = 0.$$

$$\Rightarrow \sum_{t=1}^{n}\left[\frac{x^t-(\theta-1)}{(\theta-1)^2}\right] = 0.$$

$$\Rightarrow \frac{1}{(\theta-1)^2}\times\left[\sum_{t=1}^{n}x^t - n(\theta-1)\right] = 0.$$

$\therefore$ Since, $\theta > 1$, $\sum_{t=1}^{n} x^t - n(\theta-1) = 0$

$$\Rightarrow \text{MLE } \boxed{\theta = \frac{\sum_{t=1}^{n} x^t}{n} + 1}$$

(b) $f(x|\theta) = (\theta-1).x^{\theta-2}$, $0 \le x \le 1$, $1 \le \theta < \infty$

$$L_n(\theta) = \prod_{t=1}^{n}(\theta-1).x^{t^{\theta-2}}$$

$$= (\theta-1)^n . \prod_{t=1}^{n} x^{t^{\theta-2}}.$$

$$\therefore \log(L_n(\theta)) = n\log(\theta-1) + (\theta-2)\sum_{t=1}^{n}\log x^t.$$

$$\frac{\partial(\log L_n(\theta))}{\partial\theta} = 0$$

$$\Rightarrow \frac{n}{\theta-1} + \sum_{t=1}^{n}\log x^t = 0.$$

$$\Rightarrow \text{MLE } \boxed{\theta = \frac{-n}{\sum_{t=1}^{n}\log x^t} + 1}$$

(c) $f(x|\theta) = \frac{1}{\theta}$, $0 \le x \le \theta$, $\theta > 0$.

$$L_n(\theta) = \prod_{t=1}^{n}\left(\frac{1}{\theta}\right)$$

$$L_n(\theta) = \left(\frac{1}{\theta}\right)^n.$$

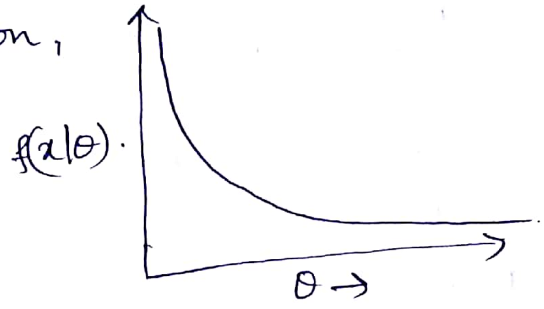
To maximize $L_n(\theta)$, we need to choose θ as low as possible.

But, $0 \leq x \leq \theta$ condition need to be satisfied.

Hence, choose small θ such that $\theta \geq \max(x_i^t)$.

$\therefore$ MLE $\boxed{\theta = \max(x^t)}$

Likelihood function,



Q3) $P(x|C)$ denote Bernoulli density function for class $C \in \{C_1, C_2\}$.
and $P(C)$ the prior.

(a) $P_1 = P(x=0|C_1)$, $P_2 = P(x=0|C_2)$, $P(C_1)$ & $P(C_2)$ are the priors.

Let evidence $= P_e$

given $x=0$

$$P(C_1|x) = \frac{P(x=0|C_1) \cdot P(C_1)}{P_e} = \frac{P_1 \times P(C_1)}{P_e}$$

$$P(C_2|x) = \frac{P(x=0|C_2) \cdot P(C_2)}{P_e} = \frac{P_2 \times P(C_2)}{P_e}$$

$\therefore$ Classification rule for $x=0$,

$$\text{choose} \begin{cases} C_1, & \text{if } P_1 \times P(C_1) > P_2 \times P(C_2) \\ C_2 & \text{otherwise} \end{cases}$$

given $x=1$

$$P(C_1|x) = \frac{P(x=1|C_1) \cdot P(C_1)}{P_e} = \frac{(1 - P(x=0|C_1)) \cdot P(C_1)}{P_e} = \frac{(1-P_1) \times P(C_1)}{P_e}$$

$$P(C_2|x) = \frac{P(x=1|C_2) \cdot P(C_2)}{P_e} = \frac{(1 - P(x=0|C_2)) \cdot P(C_2)}{P_e} = \frac{(1-P_2) \cdot P(C_2)}{P_e}$$

$\therefore$ Classification rule for $x=1$,

$$\text{choose} \begin{cases} C_1, & \text{if } (1-P_1) \cdot P(C_1) > (1-P_2) \cdot P(C_2) \\ C_2 & \text{otherwise} \end{cases}$$

(b) D-dimensional independent Bernoulli densities specified by $p_{ij} = p(x_j=0|C_i)$ for $i=1,2$ and $j=1,2---D$.

$$\therefore \quad P(x|C_i) = \prod_{j=1}^{D} p_{ij}^{(1-x_j)} \cdot (1-p_{ij})^{x_j}$$

$$\Rightarrow \quad P(x|C_1) = \prod_{j=1}^{D} p_{1j}^{(1-x_j)} \cdot (1-p_{1j})^{x_j}$$

and $$P(x|C_2) = \prod_{j=1}^{D} p_{2j}^{(1-x_j)} \cdot (1-p_{2j})^{x_j}$$

Determinant functions,

$$\therefore \quad \boxed{g_1(x) = p(x|C_1) \cdot P(C_1)}$$

$$g_1(x) = \left[\prod_{j=1}^{D} p_{1j}^{(1-x_j)} \cdot (1-p_{1j})^{x_j}\right] \cdot P(C_1)$$

It is sufficient to compare determinant functions as evidence is common for C_1 & C_2. (same)

Taking log both sides,

$$\log g_1(x) = \left[\sum_{j=1}^{D}\left[(1-x_j)\log p_{1j} + x_j \log(1-p_{1j})\right]\right] + \log(P(C_1)).$$

Similarly,

$$\boxed{g_2(x) = P(x|C_2) \cdot P(C_2)}$$

$$\therefore \quad g_2(x) = \left[\prod_{j=1}^{D} p_{2j}^{(1-x_j)} \cdot (1-p_{2j})^{x_j}\right] \cdot P(C_2).$$

Taking log both sides,

$$\log g_2(x) = \left[\sum_{j=1}^{D}\left[(1-x_j)\log p_{2j} + x_j \log(1-p_{2j})\right]\right] + \log(P(C_2)).$$

<u>Classification rule</u>:

$$\text{choose} \begin{cases} C_1, & \text{if } \log(g_1(x)) > \log(g_2(x)) \\ C_2, & \text{otherwise} \end{cases}$$

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

and $P(C_1) = 0.2, 0.6, 0.8$ with $P(C_2) = 1 - P(C_1)$.

i) for $P(C_1) = 0.2$ and $(x_1, x_2) = (0,0)$. $\Rightarrow P(C_2) = 0.8$

$$P(C_1|x) = \frac{P(x|C_1) \cdot P(C_1)}{P(x)} \quad \text{and} \quad P(C_2|x) = \frac{P(x|C_2) \cdot P(C_2)}{P(x)}.$$

$$P(x|C_1) = P(x_1=0|C_1) \cdot P(x_2=0|C_1)$$
$$= P_{11} \times P_{12} = 0.1 \times 0.6 = 0.06$$
$$P(x|C_2) = P(x_1=0|C_1) \cdot P(x_2=0|C_1)$$
$$= P_{21} \times P_{22} = 0.6 \times 0.9 = 0.54$$

But, $P(x) = P(x|C_1) \cdot P(C_1) + P(x|C_2) \cdot P(C_2)$.
$$= 0.06 \times 0.2 + 0.54 \times 0.8$$
$$= 0.444$$

$$\therefore \quad P(C_1|x) = \frac{0.06 \times 0.2}{0.444} = 0.027$$

$$P(C_2|x) = \frac{0.54 \times 0.8}{0.444} = 0.9729$$

Computations for the other 11 cases are done using matlab.
Code is in Q3c.m and output in Q3cout.txt. file.

Q3c)

Output for all the 12 cases from Matlab code:

P(C1|x) = 0.027027, P(C2|x) =0.972973 for P(C1) = 0.200000 and x = (0,0)

P(C1|x) = 0.692308, P(C2|x) =0.307692 for P(C1) = 0.200000 and x = (0,1)

P(C1|x) = 0.027027, P(C2|x) =0.972973 for P(C1) = 0.200000 and x = (1,0)

P(C1|x) = 0.692308, P(C2|x) =0.307692 for P(C1) = 0.200000 and x = (1,1)

P(C1|x) = 0.142857, P(C2|x) =0.857143 for P(C1) = 0.600000 and x = (0,0)

P(C1|x) = 0.931034, P(C2|x) =0.068966 for P(C1) = 0.600000 and x = (0,1)

P(C1|x) = 0.142857, P(C2|x) =0.857143 for P(C1) = 0.600000 and x = (1,0)

P(C1|x) = 0.931034, P(C2|x) =0.068966 for P(C1) = 0.600000 and x = (1,1)

P(C1|x) = 0.307692, P(C2|x) =0.692308 for P(C1) = 0.800000 and x = (0,0)

P(C1|x) = 0.972973, P(C2|x) =0.027027 for P(C1) = 0.800000 and x = (0,1)

P(C1|x) = 0.307692, P(C2|x) =0.692308 for P(C1) = 0.800000 and x = (1,0)

P(C1|x) = 0.972973, P(C2|x) =0.027027 for P(C1) = 0.800000 and x = (1,1)

Q4)

Table of error rate of each prior on the validation set:

Num of correct pred: 68, Error rate: 23.5955% and sigma: -5
Num of correct pred: 71, Error rate: 20.2247% and sigma: -4
Num of correct pred: 69, Error rate: 22.4719% and sigma: -3
Num of correct pred: 70, Error rate: 21.3483% and sigma: -2
Num of correct pred: 68, Error rate: 23.5955% and sigma: -1
Num of correct pred: 64, Error rate: 28.0899% and sigma: 0
Num of correct pred: 64, Error rate: 28.0899% and sigma: 1
Num of correct pred: 60, Error rate: 32.5843% and sigma: 2
Num of correct pred: 60, Error rate: 32.5843% and sigma: 3
Num of correct pred: 60, Error rate: 32.5843% and sigma: 4
Num of correct pred: 61, Error rate: 31.4607% and sigma: 5
Best performance on validation set for sigma = -4 with error rate = 20.2247%

Error rate using the best prior on the test set:

Error Rate using the best prior: 14.6067%