HW 1

Arnab Dey
Student ID: 5563169
Email: dey00011@umn.edu

Solution 1.a

I_1 is denoted by a single closed interval $[a_1, b_1] \in \mathbb{R}$. VC-dimension of I_1 is $VC(I_1) = 2$.

Proof

The single closed interval $[a_1, b_1]$ can shatter 2 points on real line as shown below:

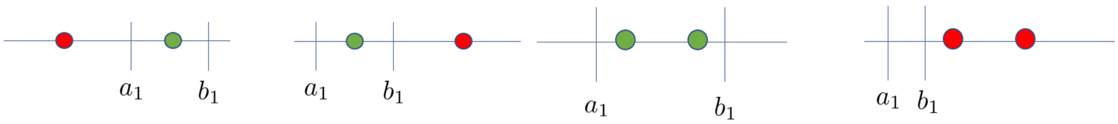

Figure 1: I_1 can shatter 2 points

But it cannot shatter 3 points on real line in the following configuration:

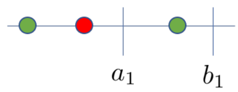

Figure 2: I_1 cannot shatter 3 points

Solution 1.b

I_2 is denoted by union of two closed intervals i.e. $[a_1, b_1] \cup [a_2, b_2] \in \mathbb{R}$. VC-dimension of I_2 is $VC(I_2) = 4$.

Proof

union of two closed intervals i.e. $[a_1, b_1] \cup [a_2, b_2]$ can shatter 4 points on real line as shown below. For clarity, all the combinations are not shown:

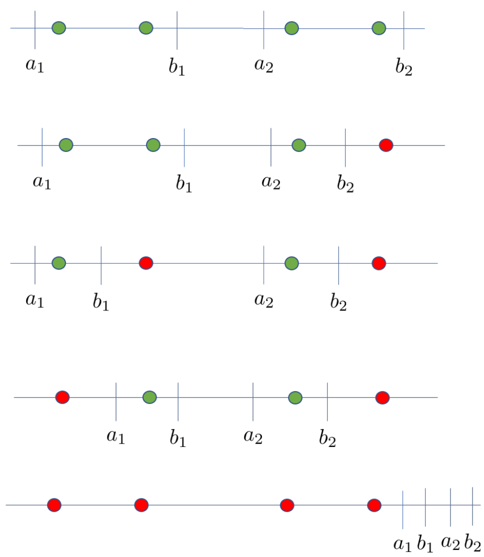

Figure 3: I_2 can shatter 4 points

But it cannot shatter 5 points on real line in the following configuration:

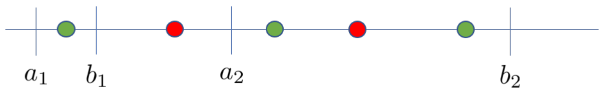

Figure 4: I_2 cannot shatter 5 points

Solution 1.c

I_K is denoted by union of K closed intervals i.e. $[a_1, b_1] \cup [a_2, b_2] \cup \ldots \cup [a_K, b_K] \in \mathbb{R}$. VC-dimension of I_K is $VC(I_K) = 2K$.

Proof

As any single closed interval can shatter 2 points, union of K such disjoint intervals will be able to shatter $2 \times K$ points. But any alternate combination of $2K + 1$ negative and positive samples cannot be shattered by union of K disjoint intervals. In case of overlapped intervals, the union can be considered as a single closed interval and therefore, the number of points it can shatter will be always less than disjoint intervals.

Solution 2.a

pdf is given by $f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, where $x > 0, \theta > 0$. Sample size is n. The set of all samples is denoted by $\{x^t\}_{t=1}^n$.

MLE Calculation

The likelihood function is given by

$$L = \prod_{t=1}^{n} \frac{1}{\theta} e^{-\frac{x^t}{\theta}}$$

Therefore, the log-likelihood function is given by

$$F = \ln L = \sum_{t=1}^{n} \left[-\ln\theta - \frac{x^t}{\theta} \right]$$

To maximize the log-likelihood, we find $\frac{\partial F}{\partial \theta}$ and set it to 0 to find the estimate of the parameter $\theta = \hat{\theta}$

$$\begin{aligned}
\frac{\partial F}{\partial \theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{t=1}^{n} x^t \\
\implies 0 &= -\frac{n}{\hat{\theta}} + \frac{1}{\hat{\theta}^2} \sum_{t=1}^{n} x^t \\
\implies \hat{\theta} &= \frac{\sum_{t=1}^{n} x^t}{n} \qquad [\because \hat{\theta} > 0]
\end{aligned}$$

Solution 2.b

pdf is given by $f(x|\theta) = 2\theta x^{2\theta-1}$, where $0 < x \leq 1, 0 < \theta < \infty$. Sample size is n. The set of all samples is denoted by $\{x^t\}_{t=1}^n$.

MLE Calculation

The likelihood function is given by

$$L = \prod_{t=1}^{n} 2\theta {x^t}^{2\theta-1}$$

Therefore, the log-likelihood function is given by

$$\begin{aligned} F = \ln L &= \sum_{t=1}^{n} \left[\ln 2 + \ln\theta + (2\theta - 1)\ln x^t\right] \\ &= n\ln 2 + n\ln\theta + (2\theta-1)\sum_{t=1}^{n}\ln x^t \end{aligned}$$

To maximize the log-likelihood, we find $\frac{\partial F}{\partial \theta}$ and set it to 0 to find the estimate of the parameter $\theta = \hat{\theta}$

$$\begin{aligned} \frac{\partial F}{\partial \theta} &= \frac{n}{\theta} + 2\sum_{t=1}^{n}\ln x^t \\ \implies 0 &= \frac{n}{\hat{\theta}} + 2\sum_{t=1}^{n}\ln x^t \\ \implies \hat{\theta} &= -\frac{n}{2\sum_{t=1}^{n}\ln x^t} \end{aligned}$$

Solution 2.c

pdf is given by $f(x|\theta) = \frac{1}{2\theta}$, where $0 \le x \le 2\theta$. Sample size is n. The set of all samples is denoted by $\{x^t\}_{t=1}^{n}$.

MLE Calculation

The likelihood function is given by

$$L = \prod_{t=1}^{n}\left[\frac{1}{2\theta}\right] = \frac{1}{(2\theta)^n}$$

As this is a decreasing function of θ, L will be maximum when θ will be minimum for a given number of samples(n). As $0 \le x \le 2\theta$, $\theta_{min} = \frac{max(x^1,x^2,...,x^n)}{2}$

Solution 3.a

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

Given

$$\begin{aligned} &P(C_1), P(C_2), \text{ and considering exhaustive events, } P(C_1) + P(C_2) = 1 \\ &p_1 = P(x=0|C_1) \implies (1-p_1) = P(x=1|C_1) \\ &p_2 = P(x=0|C_2) \implies (1-p_2) = P(x=1|C_2) \end{aligned}$$

By Bayes rule, the posteriors are given by, $P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}, i \in \{1,2\}$.
In case of Bernoulli density function, we have

$$P(x|C_i) = p_i^{1-x}(1-p_i)^x, x \in \{0,1\}, i \in \{1,2\}$$

and,

$$P(x) = \sum_{i=1}^{2} P(x|C_i)P(C_i)$$

For classification based on posteriors, we can create discriminant functions as follows,

$$g_i(x) = P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}, i \in \{1, 2\}$$

As for both the values of i, the denominator is same, we can take the decision based on $P(x|C_i)P(C_i)$. Therefore, our discriminant function can be reduced to

$$\begin{aligned} g_i(x) &= P(x|C_i)P(C_i), i \in \{1, 2\} \\ &= p_i^{1-x}(1-p_i)^x P(C_i) \end{aligned} \tag{1}$$

or equivalently, (natural logarithm of the above discriminant function),

$$g_i(x) = (1-x)\ln p_i + x\ln(1-p_i) + \ln(P(C_i)), i \in \{1, 2\} \tag{2}$$

Classification Rule Choose C_i if $g_i(x) = \max_k g_k(x)$
In other words, choose C_1 if,

$$\begin{aligned} &(1-x)\ln p_1 + x\ln(1-p_1) + \ln(P(C_1)) \geq (1-x)\ln p_2 + x\ln(1-p_2) + \ln(P(C_2)) \\ \implies &(1-x)\ln\left(\frac{p_1}{p_2}\right) \geq x\ln\left(\frac{1-p_2}{1-p_1}\right) + \ln\left(\frac{1-P(C_1)}{P(C_1)}\right) \end{aligned} \tag{3}$$

Solution 3.b

For D-dimensional independent Bernoulli densities specified by $p_{ij} = P(x_j = 0|C_i), i \in \{1, 2\}, j \in \{1, 2, \ldots, D\}$,

$$P(x|C_i) = \prod_{j=1}^{D} p_{ij}^{1-x_j}(1-p_{ij})^{x_j} \tag{4}$$

Therefore, as in case of Solution 3.a, the discriminant function can be given as follows,

$$g_i(x) = P(x|C_i)P(C_i) = \left[\prod_{j=1}^{D} p_{ij}^{1-x_j}(1-p_{ij})^{x_j}\right] P(C_i)$$

or equivalently, (natural logarithm of the above discriminant function),

$$g_i(x) = \sum_{j=1}^{D} [(1-x_j)\ln p_{ij} + x_j\ln(1-p_{ij})] + \ln(P(C_i)) \tag{5}$$

Classification Rule for D-dimensional Case Choose C_i if $g_i(x) = \max_k g_k(x)$
In other words, choose C_1 if,

$$\begin{aligned} &\sum_{j=1}^{D} [(1-x_j)\ln p_{1j} + x_j\ln(1-p_{1j})] + \ln(P(C_1)) \geq \sum_{j=1}^{D} [(1-x_j)\ln p_{2j} + x_j\ln(1-p_{2j})] + \ln(P(C_2)) \\ \implies &\sum_{j=1}^{D} [(1-x_j)\ln p_{1j} + x_j\ln(1-p_{1j})] \geq \sum_{j=1}^{D} [(1-x_j)\ln p_{2j} + x_j\ln(1-p_{2j})] + \ln\left(\frac{1-P(C_1)}{P(C_1)}\right) \end{aligned}$$

Solution 3.c

Posterior probability is given by,

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}, i \in \{1, 2\}$$

where $P(x|C_i)$ is given by (4). $P(x)$ is given by

$$P(x) = \sum_{i=1}^{2} P(x|C_i)P(C_i)$$

Posterior probabilities for different samples for different priors is tabulated below. Detail calculation has been given after the table.

Samples	$P(C_1) = 0.2$	$P(C_1) = 0.6$	$P(C_1) = 0.8$
(0,0)	$P(C_1\|x) = 0.027$ $P(C_2\|x) = 0.973$	$P(C_1\|x) = 0.143$ $P(C_2\|x) = 0.857$	$P(C_1\|x) = 0.308$ $P(C_2\|x) = 0.692$
(0,1)	$P(C_1\|x) = 0.692$ $P(C_2\|x) = 0.308$	$P(C_1\|x) = 0.931$ $P(C_2\|x) = 0.069$	$P(C_1\|x) = 0.973$ $P(C_2\|x) = 0.027$
(1,0)	$P(C_1\|x) = 0.027$ $P(C_2\|x) = 0.973$	$P(C_1\|x) = 0.143$ $P(C_2\|x) = 0.857$	$P(C_1\|x) = 0.308$ $P(C_2\|x) = 0.692$
(1,1)	$P(C_1\|x) = 0.692$ $P(C_2\|x) = 0.308$	$P(C_1\|x) = 0.931$ $P(C_2\|x) = 0.069$	$P(C_1\|x) = 0.973$ $P(C_2\|x) = 0.027$

Calculation

Sample: $\mathbf{x}$ = (0,0) From (4),

$$P(x|C_1) = p_{11}p_{12} = 0.6 \times 0.1 = 0.06$$
$$P(x|C_2) = p_{21}p_{22} = 0.6 \times 0.9 = 0.54$$

Therefore,

$$P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2) = 0.06 \times P(C_1) + 0.54 \times P(C_2) \tag{6}$$

For $P(C_1) = 0.2, P(C_2) = 0.8$,

$$\begin{aligned} P(x) &= 0.444 \\ P(C_1|x) &= \frac{0.06 \times 0.2}{0.444} \\ &= 0.027 \\ P(C_2|x) &= \frac{0.54 \times 0.8}{0.444} \\ &= 0.973 \end{aligned} \tag{7}$$

For $P(C_1) = 0.6, P(C_2) = 0.4$,

$$\begin{aligned} P(x) &= 0.252 \\ P(C_1|x) &= \frac{0.06 \times 0.6}{0.252} \\ &= 0.143 \\ P(C_2|x) &= \frac{0.54 \times 0.4}{0.252} \\ &= 0.857 \end{aligned} \tag{8}$$

For $P(C_1) = 0.8, P(C_2) = 0.2$,

$$\begin{aligned} P(x) &= 0.156 \\ P(C_1|x) &= \frac{0.06 \times 0.8}{0.156} \\ &= 0.308 \\ P(C_2|x) &= \frac{0.54 \times 0.2}{0.156} \\ &= 0.692 \end{aligned} \tag{9}$$

Sample: $\mathbf{x} = \mathbf{(0,1)}$ From (4),

$$P(x|C_1) = p_{11}(1-p_{12}) = 0.6 \times 0.9 = 0.54$$
$$P(x|C_2) = p_{21}(1-p_{22}) = 0.6 \times 0.1 = 0.06$$

Therefore,

$$P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2) = 0.54 \times P(C_1) + 0.06 \times P(C_2) \tag{10}$$

For $P(C_1) = 0.2, P(C_2) = 0.8$,

$$\begin{aligned} P(x) &= 0.156 \\ P(C_1|x) &= \frac{0.54 \times 0.2}{0.156} \\ &= 0.692 \\ P(C_2|x) &= \frac{0.06 \times 0.8}{0.156} \\ &= 0.3077 \end{aligned} \tag{11}$$

For $P(C_1) = 0.6, P(C_2) = 0.4$,

$$\begin{aligned} P(x) &= 0.348 \\ P(C_1|x) &= \frac{0.54 \times 0.6}{0.348} \\ &= 0.931 \\ P(C_2|x) &= \frac{0.06 \times 0.4}{0.348} \\ &= 0.069 \end{aligned} \tag{12}$$

For $P(C_1) = 0.8, P(C_2) = 0.2$,

$$\begin{aligned} P(x) &= 0.444 \\ P(C_1|x) &= \frac{0.54 \times 0.8}{0.444} \\ &= 0.973 \\ P(C_2|x) &= \frac{0.06 \times 0.2}{0.444} \\ &= 0.027 \end{aligned} \tag{13}$$

Sample: $\mathbf{x} = \mathbf{(1,0)}$ From (4),

$$P(x|C_1) = (1-p_{11})p_{12} = 0.4 \times 0.1 = 0.04$$
$$P(x|C_2) = (1-p_{21})p_{22} = 0.4 \times 0.9 = 0.36$$

Therefore,

$$P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2) = 0.04 \times P(C_1) + 0.36 \times P(C_2)$$

For $P(C_1) = 0.2, P(C_2) = 0.8$,

$$\begin{aligned} P(x) &= 0.296 \\ P(C_1|x) &= \frac{0.04 \times 0.2}{0.296} \\ &= 0.027 \\ P(C_2|x) &= \frac{0.36 \times 0.8}{0.296} \\ &= 0.973 \end{aligned}$$

For $P(C_1) = 0.6, P(C_2) = 0.4$,

$$\begin{aligned} P(x) &= 0.168 \\ P(C_1|x) &= \frac{0.04 \times 0.6}{0.168} \\ &= 0.143 \\ P(C_2|x) &= \frac{0.36 \times 0.4}{0.168} \\ &= 0.857 \end{aligned}$$

For $P(C_1) = 0.8, P(C_2) = 0.2$,

$$\begin{aligned} P(x) &= 0.104 \\ P(C_1|x) &= \frac{0.04 \times 0.8}{0.104} \\ &= 0.308 \\ P(C_2|x) &= \frac{0.36 \times 0.2}{0.104} \\ &= 0.692 \end{aligned}$$

Sample: $\mathbf{x} = (1,1)$ From (4),

$$\begin{aligned} P(x|C_1) &= (1 - p_{11})(1 - p_{12}) = 0.4 \times 0.9 = 0.36 \\ P(x|C_2) &= (1 - p_{21})(1 - p_{22}) = 0.4 \times 0.1 = 0.04 \end{aligned}$$

Therefore,

$$P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2) = 0.36 \times P(C_1) + 0.04 \times P(C_2)$$

For $P(C_1) = 0.2, P(C_2) = 0.8$,

$$\begin{aligned} P(x) &= 0.104 \\ P(C_1|x) &= \frac{0.36 \times 0.2}{0.104} \\ &= 0.692 \\ P(C_2|x) &= \frac{0.04 \times 0.8}{0.104} \\ &= 0.308 \end{aligned}$$

For $P(C_1) = 0.6, P(C_2) = 0.4$,

$$\begin{aligned} P(x) &= 0.232 \\ P(C_1|x) &= \frac{0.36 \times 0.6}{0.232} \\ &= 0.931 \\ P(C_2|x) &= \frac{0.04 \times 0.4}{0.232} \\ &= 0.069 \end{aligned}$$

For $P(C_1) = 0.8, P(C_2) = 0.2$,

$$\begin{aligned} P(x) &= 0.296 \\ P(C_1|x) &= \frac{0.36 \times 0.8}{0.296} \\ &= 0.973 \\ P(C_2|x) &= \frac{0.04 \times 0.2}{0.296} \\ &= 0.027 \end{aligned}$$

Solution 4

σ	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P(C_1\|\sigma)$	0.007	0.018	0.047	0.119	0.269	0.5	0.731	0.881	0.953	0.982	0.993
Error rate(%)	54	54	54	51	49	52	45	46	46	46	46

Table 1: Error-rate Vs. σ table

σ	1
$P(C_1\|\sigma)$	0.731
Error rate(%)	47.5

Table 2: Error rate on test dataset