CSCI 5521: Homework 1
Section 811 (UNITE)

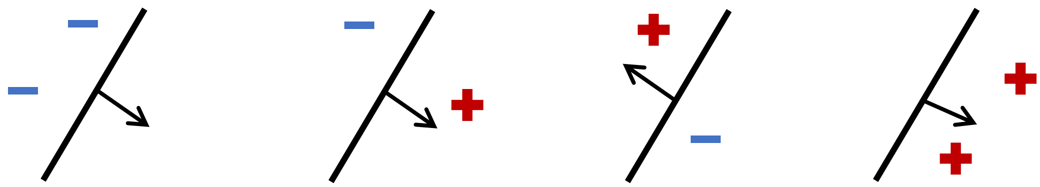
Brian Cooper (5310799)
coope824@umn.edu

Question 1a: Derive the VC dimension d_c of a threshold c in $\mathbb{R}$.

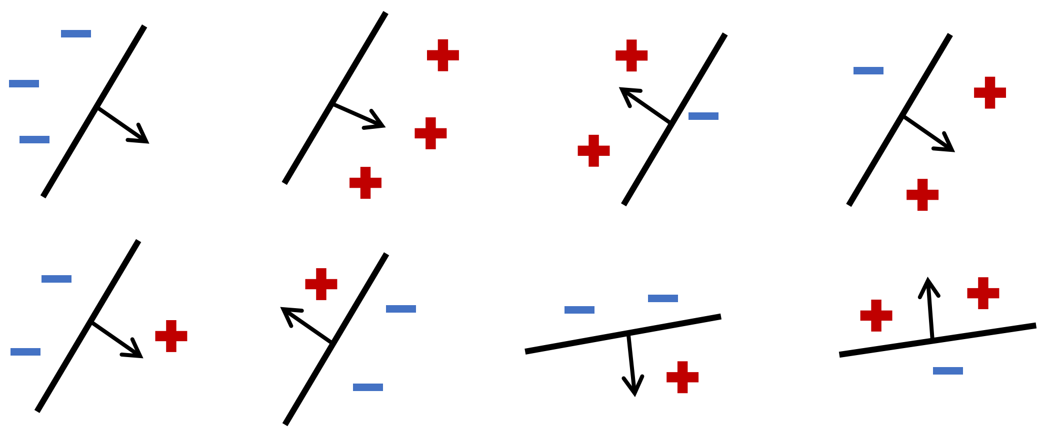
Target function $f(x) = \begin{cases} +1 & if\ x > c \\ -1 & if\ x \leq c \end{cases}$

- The target function can shatter 1 point:

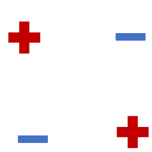
- The target function can shatter 2 points:



- The target function can shatter 3 points:



- The target function *cannot* shatter 4 points:



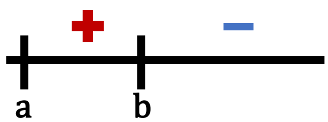
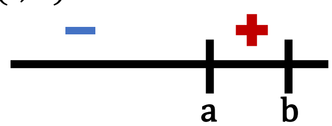
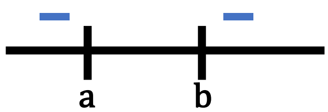
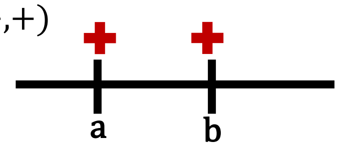
A threshold c cannot shatter this configuration of points.

Thus, the VC dimensions d_c of a threshold c in $\mathbb{R}$ is 3.

Question 1b: Derive the VC dimension d_I of intervals in $\mathbb{R}$.
Target function $[a, b]$ labels an example positive **iff** it lies in the interval

- The target function can shatter any one point, regardless of whether it is positive or negative. A positive point can fall within the interval even if a == b, and a negative point can fall outside of the interval even if a == b.

- The target function can shatter two points. The possible configurations for two points are:

Case 1. (+,-)	Case 2. (-,+)
	
Case 3. (-,-) 	Case 4. (+,+) 

- The target function cannot shatter three points. An example of this is the sequence (+, -, +):

An interval [a,b] cannot shatter this arrangement.

Thus, the VC dimensions d_I of intervals in $\mathbb{R}$ is 2.

Question 2a: Find the maximum likelihood estimation for the following pdf.

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

$$\Rightarrow \mathcal{L}_n(\theta) = \prod_{t=1}^{n} \frac{1}{\theta} e^{-\frac{x^t}{\theta}}$$

$$\Rightarrow \log[\mathcal{L}_n(\theta)] = \sum_{t=1}^{n} \log \frac{1}{\theta} e^{-\frac{x^t}{\theta}}$$

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

$$\Rightarrow \frac{\partial[\log[\mathcal{L}_n(\theta)]]}{\partial\theta} = 0$$

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

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

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

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

Thus, the MLE for $f(x|\theta)$ is $\boxed{\theta = \frac{\sum_{t=1}^{n} x^t}{n}}$

Question 2b: Find the maximum likelihood estimation for the following pdf.

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

$$\Rightarrow \mathcal{L}_n(\theta) = \prod_{t=1}^{n} 2\theta \cdot x^{t(2\theta-1)} = (2\theta)^n \prod_{t=1}^{n} x^{t(2\theta-1)}$$

$$\therefore \log[\mathcal{L}_n(\theta)] = \mathrm{n} \cdot \log(2\theta) + (2\theta - 1)\sum_{t=1}^{n} \log x^t$$

$$\Rightarrow \frac{\partial[\log[\mathcal{L}_n(\theta)]]}{\partial\theta} = 0$$

$$\Rightarrow \frac{n}{2\theta} + \sum_{t=1}^{n} \log x^t = 0$$

Thus, the MLE for $f(x|\theta)$ is $\boxed{\theta = -\frac{n}{\sum_{t=1}^{n} \log x^t}}$

Question 2c: Find the maximum likelihood estimation for the following pdf.

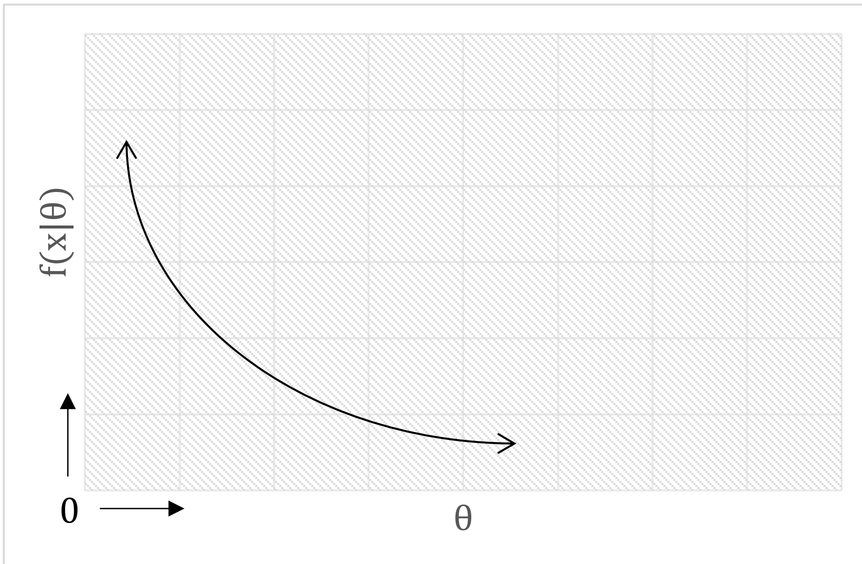
$f(x|\theta) = \frac{1}{2\theta}, \qquad 0 \leq x \leq 2\theta$

$$\Rightarrow \mathcal{L}_n(\theta) = \prod_{t=1}^{n} \frac{1}{2\theta}$$

$$= \left(\frac{1}{2\theta}\right)^2$$

Brian Cooper (5310799)
coope824@umn.edu

To maximize $\mathcal{L}_n$, minimize θ. Since 2θ is on the upper bound of x, θ must be chosen such that $2\theta \geq \max(x^t)$.



From this, we get $2\theta = \max(x^t)$, which is the equivalent to $\theta = \frac{\max(x^t)}{2}$.

Thus, the MLE for $f(x|\theta)$ is $\boxed{\theta = \frac{\max(x^t)}{2}}$

Question 3a

Bernoulli density function $P(x|c)$ for class $C\epsilon\{C_1, C_2\}$
Prior: $P(C)$, specifically $P(C_1), P(C_2)$
$p_1 \equiv p(x = 0|C_1), p_2 \equiv p(x = 0|C_2)$
Posteriors: $P(C_1|x), P(C_2|x)$
Let p_e be the evidence.

For x=0:

$$P(C_1|x) = \frac{p(x=0|C_1)\cdot P(C_1)}{p_e} = \frac{P_1 P(C_1)}{p_e}$$
$$P(C_2|x) = \frac{p(x=0|C_2)\cdot P(C_2)}{p_e} = \frac{P_2 P(C_2)}{p_e}$$

For x=1:

$$P(C_1|x) = \frac{p(x=1|C_1)\cdot P(C_1)}{p_e} = \frac{[1-p(x=0|C_1)]P(C_1)}{p_e} = \frac{(1-p_1)\cdot P(C_1)}{p_e}$$
$$P(C_2|x) = \frac{p(x=1|C_2)\cdot P(C_2)}{p_e} = \frac{[1-p(x=0|C_2)]P(C_2)}{p_e} = \frac{(1-p_2)\cdot P(C_2)}{p_e}$$

Brian Cooper (5310799)
coope824@umn.edu

Rules for classifying x = 0:

$$\begin{cases} choose\ C_1 & if\ p_1 \cdot P(C_1) > p_2 \cdot P(C_2) \\ choose\ C_2 & if\ p_1 \cdot P(C_1) \le p_2 \cdot P(C_2) \end{cases}$$

Rules for classifying x = 1:

$$\begin{cases} choose\ C_1 & if\ (1-p_1) \cdot P(C_1) > (1-p_2) \cdot P(C_2) \\ choose\ C_2 & if\ (1-p_1) \cdot P(C_1) \le (1-p_2) \cdot P(C_2) \end{cases}$$

Question 3b

$p_{ij} \equiv p\left(x_j = 0 \middle| C_i\right)\ for\ i = 1,2\ and\ j = 1,2,\ldots,D$

$$p(x|C_1) = \prod_{j=1}^{D} p_{1j}^{1-x_j} \cdot \left(1 - p_{1j}\right)^{x_j}$$

$$p(x|C_2) = \prod_{j=1}^{D} p_{2j}$$

$$f_1(x) = p(x|C_1) \cdot P(C_1)$$

$$\log[f_1(x)] = \left[\sum_{j=1}^{D}\left[\left(1-x_j\right)\log\left(p_{1j}\right) + x_j \log\left(1-p_{1j}\right)\right]\right] + \log P(C_1)$$

$$f_2(x) = p(x|C_2) \cdot P(C_2)$$

$$\log[f_2(x)] = \left[\sum_{j=1}^{D}\left[\left(1-x_j\right)\log\left(p_{2j}\right) + x_j \log\left(1-p_{2j}\right)\right]\right] + \log P(C_2)$$

Thus, $\begin{cases} choose\ C_1 & if\ \log[g_1(x)] > \log[g_2(x)] \\ choose\ C_2 & if\ \log[g_1(x)] \le \log[g_2(x)] \end{cases}$

Question 3c
$D = 2, p_{11} = 0.6,\ p_{12} = 0.1,\ p_{21} = 0.6,\ p_{22} = 0.9$
$P(C_1) = 0.2, 0.6, 0.8, P(C_2) = 1 - P(C_1)$

Case 1: $\boldsymbol{P(C_1) = 0.2, (x_1, x_2) = (0, 0)}$
$P(C_2) = 1 - P(C_1) = 0.8$

$$P(C_1|x) = \frac{P(x|C_1) \cdot P(C_1)}{P(x)}, P(C_2|x) = \frac{P(x|C_2) \cdot P(C_2)}{P(x)}$$

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

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

$\therefore P(C_1|x) = \frac{(0.06)(0.2)}{0.444} = 0.027$ and $P(C_2|x) = \frac{(0.54)(0.8)}{0.444} = 0.9729$

There are 11 more cases, and they all have similar logic. The results are shown below:

Case 2: $\boldsymbol{P(C_1) = 0.2, (x_1, x_2) = (0, 1)}$
$P(C_1|x) = 0.6923,\ P(C_2|x) = 0.3077$

Case 3: $\boldsymbol{P(C_1) = 0.2, (x_1, x_2) = (1, 0)}$
$P(C_1|x) = 0.027,\ P(C_2|x) = 0.973$

Case 4: $\boldsymbol{P(C_1) = 0.2, (x_1, x_2) = (1, 1)}$
$P(C_1|x) = 0.6923,\ P(C_2|x) = 0.3077$

Case 5: $\boldsymbol{P(C_1) = 0.6, (x_1, x_2) = (0, 0)}$
$P(C_1|x) = 0.1423,\ P(C_2|x) = 0.8571$

Case 6: $\boldsymbol{P(C_1) = 0.6, (x_1, x_2) = (0, 1)}$
$P(C_1|x) = 0.931,\ P(C_2|x) = 0.069$

Case 7: $\boldsymbol{P(C_1) = 0.6, (x_1, x_2) = (1, 0)}$
$P(C_1|x) = 0.1429,\ P(C_2|x) = 0.8571$

Case 8: $\boldsymbol{P(C_1) = 0.6, (x_1, x_2) = (1, 1)}$
$P(C_1|x) = 0.931,\ P(C_2|x) = 0.069$

Case 9: $\boldsymbol{P(C_1) = 0.8, (x_1, x_2) = (0, 0)}$
$P(C_1|x) = 0.3077,\ P(C_2|x) = 0.6923$

Case 10: $\boldsymbol{P(C_1) = 0.8, (x_1, x_2) = (0, 1)}$
$P(C_1|x) = 0.973,\ P(C_2|x) = 0.027$

Case 11: $\boldsymbol{P(C_1) = 0.8, (x_1, x_2) = (1, 0)}$
$P(C_1|x) = 0.3077,\ P(C_2|x) = 0.6923$

Case 12: $\boldsymbol{P(C_1) = 0.8, (x_1, x_2) = (1, 1)}$
$P(C_1|x) = 0.973,\ P(C_2|x) = 0.027$

Question 4

Validation Set Error Rate (MATLAB Output from Bayes_Learning.m)

For sigma = 1.000000e-05, Number of correct predictions: 92, Error rate: 54.00%
For sigma = 1.000000e-04, Number of correct predictions: 92, Error rate: 54.00%
For sigma = 1.000000e-03, Number of correct predictions: 92, Error rate: 54.00%
For sigma = 1.000000e-02, Number of correct predictions: 92, Error rate: 54.00%
For sigma = 1.000000e-01, Number of correct predictions: 98, Error rate: 51.00%
For sigma = 1, Number of correct predictions: 97, Error rate: 51.50%
For sigma = 2, Number of correct predictions: 109, Error rate: 45.50%
For sigma = 3, Number of correct predictions: 108, Error rate: 46.00%
For sigma = 4, Number of correct predictions: 108, Error rate: 46.00%
For sigma = 5, Number of correct predictions: 108, Error rate: 46.00%
For sigma = 6, Number of correct predictions: 108, Error rate: 46.00%

Best performance with sigma = 2, Error rate: 45.50%

Test Set Error Rate (MATLAB Output from Bayes_Testing.m)

Error rate using the optimal prior: 44.50%