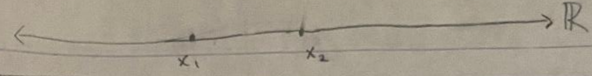
any $a \le x_1$

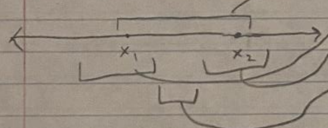
(1)



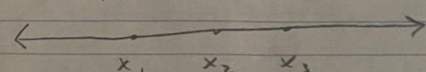
$\boxed{\text{VC dim}, d_I = 2}$

for $d_I = 1$
- can shatter one point by choosing an interval $[a, b]$ that does or does not contain the point

for $d_I = 2$ (d_I)☆
- can shatter 2 points (x_1, x_2) by choosing an interval $[a, b]$ that includes either:
  - both points
  - just the point x_1
  - just point x_2
  - neither point



for $d_I = 3$: $(d_I + 1)$☆
- cannot shatter 3 points (x_1, x_2, x_3) when $x_1 < x_2 < x_3$ given the specific example of x_1 + x_3 both being positive, with x_2 not being positive



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

likelihood $L(\theta|x) = \prod \frac{1}{\theta-1} e^{-\frac{x}{\theta-1}}$

$-\frac{x}{\theta-1}\ln(e)$

natural log

$\theta = \frac{n}{\sum x_i + n}$

$\ln(L(\theta|x)) = \sum \ln\left(\frac{1}{\theta-1}\right) + \ln\left(e^{\frac{-x}{\theta-1}}\right)$

$= \sum \left(\ln\left(\frac{1}{\theta-1}\right) - \frac{x}{\theta-1}\right)$

derivative w/ respect to θ

set = 0

$\frac{\partial}{\partial\theta} \ln L(\theta|x) = -\sum \frac{1}{\theta-1} + \sum \frac{x}{(\theta-1)^2}$

$0 = -\sum \frac{1}{\theta-1} + \sum \frac{x}{(\theta-1)^2}$

mult by $(\theta-1)^2$

$0 = -\sum(\theta-1) + \sum x_i$

$n\theta - n = \sum x_i$

$\theta = \frac{\sum x_i + N}{N} = \boxed{\frac{\sum x_i}{N} + 1}$

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

Log likelihood = $L(\theta|x) = \prod (\theta-1)x^{\theta-2}$ $\ln(\theta-1) + \ln(x^{\theta-2})$

natural log

$\ln(\theta$

$$= \sum \left[\ln(\theta-1) + (\theta-2)\ln(x) \right]$$

derive w/ respect to θ

set to 0

$$\frac{\partial L(\theta)}{\partial \theta} = \sum \left[\frac{1}{\theta-1} + \ln(x) \right] = 0$$

multiply all by $\theta - 1$

$$\sum 1 + \sum(\theta-1)(\ln(x)) = 0$$

$$N + (\theta-1)\sum \ln(x)$$

$(\theta-1)\sum \ln(x) = -N$

$\div \sum \ln(x)$

$$\theta - 1 = \frac{-N}{\sum \ln(x)}$$

$$\boxed{\theta = 1 - \frac{N}{\sum \ln(x)}}$$

2(c)

$\frac{1}{\theta}$, $0 < x \leq \theta$

pdf of observing all n data points

$L(\theta|x) = \prod \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^n$

likelihood func decreases as θ increases

so max likelihood is at smallest possible θ that satisfies parameters

$[\theta \geq \max\{x_1, x_2 \ldots x_n\}]$

hence the MLE of θ is the max value of observed sample so

$\theta = \max\{x_1, x_2 \ldots x_n\}$

(3)

density func: $P(x|C)$

$C \in \{C_1, C_2\}$

Prior : $P(C)$

a) densities $p_1 \equiv p(x=0|C_1)$, $p_2 \equiv p(x=1|C_2)$

$$P(c_1|x) = \frac{P(x|C_1)\,P(C_1)}{P(x)} \qquad P(c_2|x) = \frac{P(x|C_2)\,P(C_2)}{P(x)}$$

$$P(x) = P(x|C_1)\cdot P(C_1) + P(x|C_2)\,P(C_2)$$

~~Since assume P(x)~~

Since density func is used for both C_1 & C_2

$x = 0$

if $P(c_1|x) > P(c_2|x)$ classify x as $C1$

if " $<$ " classify as C_2

if " $=$ " classify as either

$x=0$ so for bernoulli's

	classif.
$p_1(1-P(C_1)) > p_2(1-P(C_2))$	$C1$
" $<$ "	$C2$
" $=$ "	either

$x = 1$

	classify
if $P(c_1|x) > P(c_2|x)$	$C1$
" $<$ "	C_2
" $=$ "	either

$x=1$ Bernoullis

	class
$(1-p_1)(1-P(C_1)) > (1-p_2)(1-P(C_2))$	$C1$
" $<$ "	C_2
" $=$ "	eith

3 b ~~probability of observing x given~~

$$P(x|C) = P(x_1|C) * P(x_2|C) \ldots * P(x_D|C)$$

where $P(x_j|C) = p_{ij}$ if $x_j = 0$ & $P(x_j|C) = 1 - p_{ij}$ if $x_j = 1$

Bayes rule

$$P(C|x) = \frac{P(x|C)\,P(C)}{P(x)}$$

$$P(x) = P(x|C_1)\,P(C1) + P(x|C_2)\,P(C_2)$$

x is classified as C1 if $P(C_1|x) > P(C_2|x)$,

and classified as C2 otherwise

$$P(C1|x) = \frac{P(C1)\prod(1 - p_{ij}x_j)}{P(C1)\prod(1 - p_{ij}x_j) + P(C2)\prod(1 - p_{ij}x_j)}$$

$$P(C_2|x) = \frac{P(C2)\prod(1 - p_{ij}x_j)}{P(C1)\prod(1 - p_{ij}x_j) + P(C2)\prod(1 - p_{ij}x_j)}$$

$p_{ij}x_j = p_{ij}$ when $x_j = 0$

$p_{ij}x_j = 1 - p_{ij}$ when $x_j = 1$

5

3c:

x = (0, 0): P(C1|x) = nan, P(C2|x) = nan

x = (0, 1): P(C1|x) = nan, P(C2|x) = nan

x = (1, 0): P(C1|x) = nan, P(C2|x) = nan

x = (1, 1): P(C1|x) = 0.2, P(C2|x) = 0.8

Priors: P(C1) = 0.6, P(C2) = 0.4

x = (0, 0): P(C1|x) = nan, P(C2|x) = nan

x = (0, 1): P(C1|x) = nan, P(C2|x) = nan

x = (1, 0): P(C1|x) = nan, P(C2|x) = nan

x = (1, 1): P(C1|x) = 0.6, P(C2|x) = 0.4000000000000001

Priors: P(C1) = 0.8, P(C2) = 0.19999999999999996

x = (0, 0): P(C1|x) = nan, P(C2|x) = nan

x = (0, 1): P(C1|x) = nan, P(C2|x) = nan

x = (1, 0): P(C1|x) = nan, P(C2|x) = nan

x = (1, 1): P(C1|x) = 0.8, P(C2|x) = 0.19999999999999993

[see code problem3c.py]

4: Table of Error Rates:

Prior Value	Error Rate
0.000010	0.540000
0.000100	
0.540000	
0.001000	0.540000
0.010000	0.540000
0.100000	0.510000
1.000000	0.515000
2.000000	0.455000
3.000000	0.460000
4.000000	0.460000
5.000000	0.460000
6.000000	0.460000

The error rate with the best prior on the test is: 0.44499999999999995