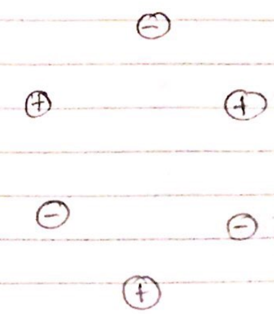
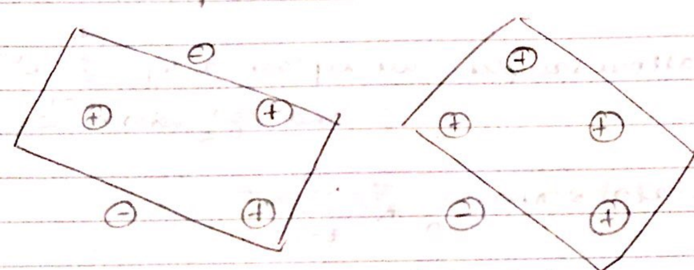
Q1(a). H is a space of all rectangles. It cannot shatter 6 points. The v.c dimension is 5.



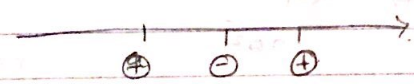
There is no rectangle to shatter this configuration.

It can shatter 5 points



(b) vc dimension of ,dz, of intervals in R.

It shatters 2 points. Its v.c. dimension is 2. It cannot shatter 3 points



→ cannot be shattered by an interval.

Q2.

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

Taking log of $f(x|\theta)$

$\Rightarrow \log \frac{1}{\theta} e^{-\frac{x}{\theta}}$.

$= \log \frac{1}{\theta} - \frac{x}{\theta} \log e$

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

for all our samples we replace x by $\sum_{t=1}^{N} x^t$

and $\log \frac{1}{\theta}$ with $n \log \frac{1}{\theta}$

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

differentiating w.r.t θ.

$\frac{\partial L(x|\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum_{t=1}^{n} x^t}{n\theta^2} = 0.$

$\cancel{\sum x} \quad -\frac{n}{\cancel{\theta}} = \frac{\sum x^t}{n\theta^{\cancel{2}}}$

$\theta = \frac{\sum_{t=1}^{n} x^t}{n^2}$

$f(x|\theta) = \theta x^{\theta-1}$

Taking log of $f(x|\theta)$

$$= \log(\theta x^{\theta-1})$$

$$L(x|\theta) = \log\theta + (\theta-1)\log x$$

$$L(x|\theta) = \log\theta + \sum_t \theta \log x^t - \sum \log x^t$$

$$\frac{\partial L(x|\theta)}{\partial\theta} = \frac{1}{\theta} + \sum_t \log x^t - 0 = 0.$$

$$\theta = \frac{1}{-\sum_t \log x^t} =$$

Since all of x^t lies between $[0,1]$, therefore the summation is a product since $\sum \log(A) + \log(B) = \log(ab)$.

all of $-\sum_t \log x^t$ will be and -ve and thus the product will be a +ve term thus allowing θ to lie in $[\theta, \infty]$.

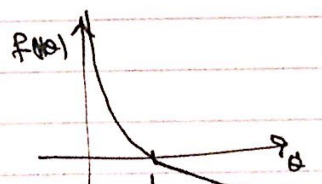
$f(x|\theta) = \frac{1}{\theta}.$

Taking log.

$\mathcal{L}(x|\theta) = \log(1/\theta). = -\log\theta.$

~~since.~~

The plot of log $f(x|\theta) =$ is



The maximum θ will be defined for x data points whose parameters are plotted on this graph. The function. ~~will be~~ is independent of x. as x is limited to lie between 0 to θ and $\theta > 0$. But for $\theta > 1$ the value of $f(x|\theta)$ decreases. Thus, θ will lie between ~~θ~~ 0 and 1 and so will x. θ can then be defined as the larger value attained by x.

Q3a Since the two class covariances are considered as independent, we must use the most generalized formula for P(C|x) in the discriminant function. The formula for which is given as

$$P(X|C_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right]$$

Q3b Since S1 and S2 are the same and learned from the same data, we calculate the common variance by adding the product of the class variance with the probability of the class itself. The formula is only affected by a common Σ instead of Σ_i.

Q3c Since S1 and S2 are diagonal, i.e the parameters of the observed values are independent of one another, the matrix is reduced to a diagonal matrix that only contains the squared variances of the parameters themselves. This changes Σ to a diagonal matrix.

Q3d Since S1 and S2 are diagonal and we wish to represent $S_i = \alpha_i \boldsymbol{I}$. Here alpha is the shared variance for all the values in xi. The shared variance is the eigen value for the given sample matrix. This eigen value vector is multipled with a identity matrix of size d which is then used in the formula.