CSCI 5521 INTRO TO MACHINE LEARNING

HOMEWORK - 1

UJVAL BANGALORE UMESH banga038@umn.edu

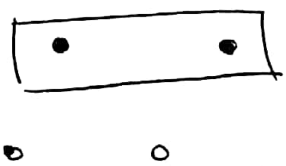
5202540

1)

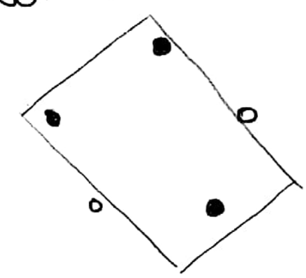
(a) Let H be the space of all rectangles.

The rectangles that are parallel to the axis and the ones that are not parallel to the axis.

The VC dimension of the parallel rectangles is 4.

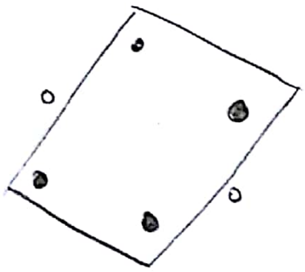


4 points can be shattered by rectangles which are parallel to the axes.



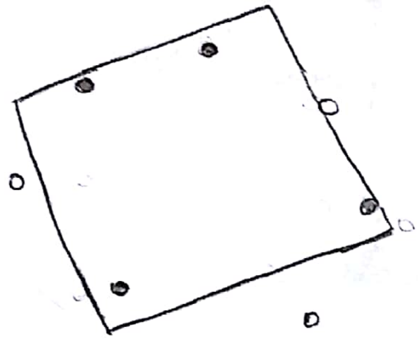
5 points can only be shattered by non-axis rectangle.

Let us now consider 6 points.



A Rotatable Rectangle can shatter 6 points.

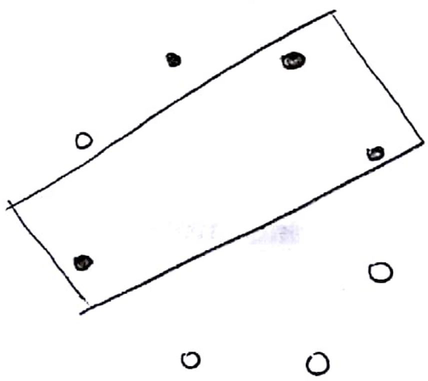
consideration of 7 points.



Rotatable rectangles can shatter 7 points.

To prove it can shatter n points, we need prove that it cannot shatter $n+1$ points.

consider 8 points.

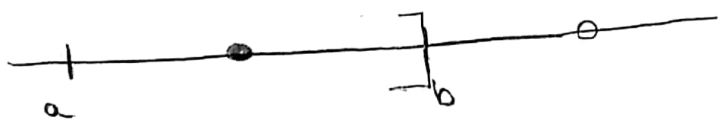


We can see that it cannot shatter 8 points.

Hence the VC dimension of rotatable rectangle is $\boxed{7.}$

b) Let d_I be the VC dimension, of intervals in in $\mathbb{R}$. The interval [a,b]

consider the interval [a,b] having just one point in it & ~~one~~ ~~outside~~ one outside



One point can be shattered.

Let us consider a positive point in between two negative points.

To shatter this arrangement, in the interval [a,b].

The VC dimension is two [2].

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

$l(\theta|\mathcal{X}) = p(\mathcal{X}|\theta) = \prod_t p(x^t|\theta)$

$L(\theta|\mathcal{X}) = \log l(\theta|\mathcal{X})$

$= \sum_t \log \left(\frac{1}{\theta} e^{-x^t/\theta} \right)$

$= \sum_t \log\left(\frac{1}{\theta}\right) + \sum_t \log \left(e^{-x^t/\theta}\right)$

$\frac{\partial}{\partial\theta} \left[\sum_t \log\left(\frac{1}{\theta}\right) + \sum_t \frac{-x^t}{\theta} \right] = 0$

$-\frac{N\theta}{\theta^2} + \sum_t \frac{x^t}{\theta^2} = 0$

$\frac{N\theta}{\theta^2} = \sum_t \frac{x^t}{\theta^2}$

$\theta = \frac{1}{N} \sum_t x^t$, $\theta > 0$

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

$L(\theta|\mathcal{X}) = \log l(\theta|\mathcal{X})$

$= \sum_t \log p(x^t|\theta)$

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

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

Scanned by CamScanner

$$= \frac{\partial}{\partial \theta}\left[\sum_t \log\theta + \sum_t (\theta-1)\log x \right]$$

$$= \frac{N}{\theta} + \sum_t (\log x)(\theta-1) = 0$$

$$\frac{N}{\theta} + \sum_{t=1}^{N} \log x = 0$$

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

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

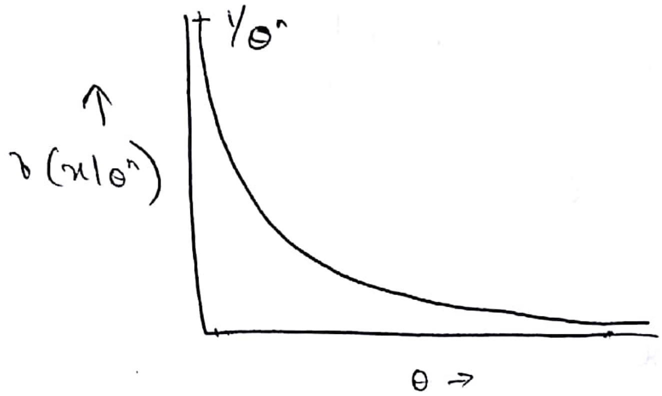
The probability density function $f_n(x|\theta)$ of $x_1, \ldots, x_n$ has the form

$$f_n(x|\theta) = \frac{1}{\theta^n}$$

MLE of θ must be a value of θ for which $\theta \ge x_i$ for $i = 1, \ldots, n$. and which maximises $1/\theta^n$. Since it is a decreasing function, the smallest θ such that $\theta \ge x_i$ for $i = 1, \ldots, n$, the estimate will be the smallest value of θ such that $\theta \ge x_i$. Since this value is $\theta = \max[x_1, \ldots, x_n]$.

Hence the MLE of θ is $\max[x_1, \ldots, x_n]$.

The function of $\frac{1}{\theta^n}$ is



~~The values of θ ...~~

3 (e)

$$P(x/C_i) = \frac{1}{(2\pi)^{d/2} |(\Sigma_i)|^{1/2}} \exp\left\{ \frac{-1}{2} (x-\mu)^T \Sigma_i^{-1} (x-\mu) \right\}$$

Estimating the loglikelihood for means.

$$L(X/\mu, \Sigma_i) = \sum_{t=1}^{N} -\frac{d}{2} \log 2\pi - \frac{1}{2} \log|\Sigma_i| - \frac{1}{2} (x-\mu)^T \Sigma_i^{-1} (x^t - \mu)$$

$$\frac{\partial L}{\partial \mu} = 0$$

$$\sum_{t=1}^{N} (x^t - \mu) = 0 \qquad N\mu = \sum_{t=1}^{N} x^t \Rightarrow \mu = \sum_{t=1}^{N} \frac{x^t}{N}$$

This can be extrapolated to $\mu_i = \sum_{t=1}^{N} \frac{x^t}{N}$

Means of different classes.

Estimating log likelihood with respect to covariance

$$\frac{\partial L}{\partial \Sigma} = 0$$

$$\frac{1}{2}\Sigma = \frac{1}{2}\sum_{t=1}^{N}(x^t-\mu)^T(x^t-\mu)$$

$$\frac{\partial}{\partial u}[a^T A a] = (A^T a)$$

$$\Sigma = \frac{\sum_{t=1}^{N}(x^t-\mu)^T(x^t-\mu)}{N}$$

$$\Sigma_i = \frac{(x^t-\mu_i)^T(x^t-\mu_i)}{N}.$$

(c) $$P(C_i|x) = \frac{P(x|C_i) \times P(C_i)}{P(x)}$$

Bayes formula.

If $S_1 = S_2$, Sigma $S = 0.6 S_1 + 0.4 S_2$.

Calculation is done based on this.

3)

d) S_1 and S_2 are diagonal.

$S_1 = \alpha_1 I$ and $S_2 = \alpha_2 I$

$S_1 - \alpha_1 I = 0$ $\quad\quad$ $S_2 - \alpha_2 I = 0$

$$\begin{bmatrix} s_{11}-\alpha_1 & & & \\ & s_{22}-\alpha_1 & & \\ & & \ddots & \\ & & & s_{nn}-\alpha_1 \end{bmatrix} = 0 \quad \begin{bmatrix} s_{11}-\alpha_2 & & & \\ & s_{22}-\alpha_2 & & \\ & & \ddots & \\ & & & s_{nn}-\alpha_2 \end{bmatrix} = 0$$

Hence α_1 and α_2 are eigen values of S_1 and S_2 respectively. We could use the eig function in matlab to obtain S_1 and S_2.

[V1, D1] = eig (S1) & [V2 D2] = eig (S2)