Notes

Q.1) $M_1 \Rightarrow$ Alice is working
$M_2 \rightarrow$ Bob is working

$Y \rightarrow$ Collected Data/tickets $\Rightarrow$ $Y_1 = 12, Y_2 = 10,$ $Y_3 = 11, Y_4 = 4, Y_5 = 11$

$$\text{Post. odds} = \frac{P(M_1|Y)}{P(M_2|Y)} = \frac{P(Y|M_1)P(M_1)}{P(Y|M_2)P(M_2)}$$

$$P(Y|M_1 = x) = e^{-10}\frac{10^x}{x!} \quad (\lambda_A = 10/hr)$$

$$P(Y|M_2 = x) = e^{-15}\frac{15^x}{x!} \quad (\lambda_B = 15/hr)$$

~~So, $P(Y|M_1) = P(Y_1|M_1) \ldots P(Y_5|M_5)$~~

$$P(Y|M_1) = P(Y|M_1 = Y_1) \ldots P(Y|M_1 = Y_5)$$

$$\Rightarrow P(Y|M_1) = \frac{e^{-50}\, 10^{12+10+11+4+11}}{12!\,10!\,11!\,4!\,11!}$$

$$\Rightarrow P(Y|M_1) = \frac{e^{-50}\, 10^{48}}{12!\,10!\,11!\,4!\,11!}$$

Teacher's Signature ____________

$P(M1) = \frac{1}{10+1} = \frac{1}{11}$, $P(M2) = 10/11$

Notes

Date / /

Page No

Similarly, $P(Y|M_2) = e^{-75} \frac{15^{48}}{12!\,11!\,10!\,4!\,\ldots\,11!}$

$$\text{Posterior odds} = \frac{P(Y|M1)\,P(M1)}{P(Y|M2)\,P(M2)}$$

$$= \frac{e^{-50}\,\frac{10^{48}}{12!\ldots 11!}}{e^{-75}\,\frac{15^{48}}{12!\ldots 11!}} \times \frac{1}{10}$$

$$= e^{25} \times \left(\frac{10}{15}\right)^{48} \times \frac{1}{10}$$

$\Rightarrow$ Ans = 25.40865

Makes Sense Given the Data,
Since Mean Data = 9.69
All points < 15.

Notes

Date / /

Page No

Q.2)

$n \to$ Noise $\sim N(0, 4)$,

$x \to$ Signal Received $\sim N(\theta, 4)$,

$\theta \sim N(5, 9)$

(Signal Transferred)

(a) Hyp. θ	Prior $f(\theta)\,d\theta$	likelihood $f(x\mid\theta)\,dx$	Bayes Num. $f(x\mid\theta)f(\theta)\,d\theta\,dx$

$$f(\theta\mid x)\,d\theta = \frac{f(x\mid\theta)\,f(\theta)\,d\theta\,dx}{f(x)}$$

$$= \frac{f(x\mid\theta)\,f(\theta)\,d\theta}{f(x)}$$

$$f(x)\,dx = \left(\int f(x\mid\theta)\,f(\theta)\,d\theta\right)dx$$

Prior $\Rightarrow$ $f(\theta) = \frac{1}{3\sqrt{2\pi}}\, e^{-\frac{(\theta-5)^2}{18}}$

Likelihood $\Rightarrow$ $f(x=6\mid\theta) = \frac{1}{2\sqrt{2\pi}}\, e^{-\frac{(6-\theta)^2}{8}}$

$\frac{74}{13}$

$f\theta$

let = N·Prior·likelihood $= \frac{1}{12\pi}\, e^{-\frac{(4\theta^2+100-40\theta+9\theta^2-108\theta+324)}{72}}$

$$= \frac{1}{12\pi}\, e^{-\left(\frac{13\theta^2-148\theta+424}{72}\right)}$$

$$= \frac{1}{12\pi}\, e^{-\frac{13}{72}\left(\theta^2-\frac{148}{13}\theta+\left(\frac{74}{13}\right)^2+\left(\frac{424}{13}-\left(\frac{74}{13}\right)^2\right)\right)}$$

$$= \frac{1}{12\pi} \times e^{-\frac{13}{72}\left(\frac{424}{13} - \frac{5476}{169}\right)} \cdot e^{-\frac{13}{72}\left(\theta - \frac{74}{13}\right)^2}$$

$$= \frac{1}{12\pi} \times e^{-\frac{1}{26}} \times$$

$$\text{Bayes Num} = N \cdot d\theta = f(x=6|\theta) f(\theta) \cdot d\theta$$

$$= \frac{e^{-1/26}}{12\pi} \cdot e^{-\frac{13}{72}\left(\theta - \frac{74}{13}\right)^2} d\theta$$

$$f(x=6) = \int_{-\infty}^{\infty} f(x=6|\theta) f(\theta) \cdot d\theta$$

$$= \frac{e^{-1/26}}{12\pi} \int_{-\infty}^{\infty} e^{-\frac{13}{72}\left(\theta - \frac{74}{13}\right)^2} d\theta$$

$$\int_{-\infty}^{\infty} f \cdot d\theta$$

Normal

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \frac{e^{-1/26}}{12\pi} \times \sqrt{2\pi} \times \frac{6}{\sqrt{13}}$$

$$=$$

Notes

$f(\theta|x)$

$$= \frac{\text{Bayes Num}}{f(x)}$$

$$= \frac{\cancel{\frac{e^{-1/26}}{\sqrt{2\pi}}} e^{-\frac{13}{72}\left(\theta-\frac{74}{13}\right)^2} \cdot d\theta}{\cancel{\frac{e^{-1/26}}{\sqrt{2\pi}}} \times \sqrt{2\pi} \times \frac{6}{\sqrt{13}}}$$

$$f(\theta|x) = \frac{1}{\sqrt{2\pi} \times \frac{6}{\sqrt{13}}} \times e^{-\frac{\left(\theta-\frac{74}{13}\right)^2}{2 \times \frac{36}{13}}} \cdot d\theta$$

$$\Downarrow$$

Posterior Pdf $\Rightarrow N\left(\frac{74}{13}, \frac{36}{13}\right)$

(b) Same $\theta \Rightarrow x_1, \ldots x_n \cdot (\bar{X} = \text{Sample Mean})$

prior $\longrightarrow$ post (From Formula)

prior $\rightarrow N(\mu_{\text{prior}}, \sigma^2_{\text{prior}})$ (θ)

Likehood $\rightarrow N(\mu, \sigma^2)$ $(x|\theta)$

Posterior $\rightarrow N(\mu_{\text{post}}, \sigma^2_{\text{post}})$ $(\theta|x)$

Teacher's Signature

$(\sigma^2 \to \sigma^2_{likel.},\ \mu \to \text{~ likhlood})$

Notes

$\theta \to 4$ times $(n=4)$

x_1, x_2, x_3, x_4 $(\bar{x}=6)$

prior $(\theta) \sim N(5, 9)$

likehood $(x|\theta) \sim N(\theta, 4)$

$(\theta|x) = ?$

$$a = \frac{1}{\sigma^2_{prior}} = \frac{1}{9}$$

$$b = \frac{n}{\sigma^2} = \frac{4}{4} = 1$$

$$\mu_{post} = \frac{a\mu_{prior} + b\bar{x}}{a+b}$$

$$= \frac{5/9 + 6}{\frac{10}{9}} = \frac{\frac{59}{9}}{\frac{10}{9}} = 5.9$$

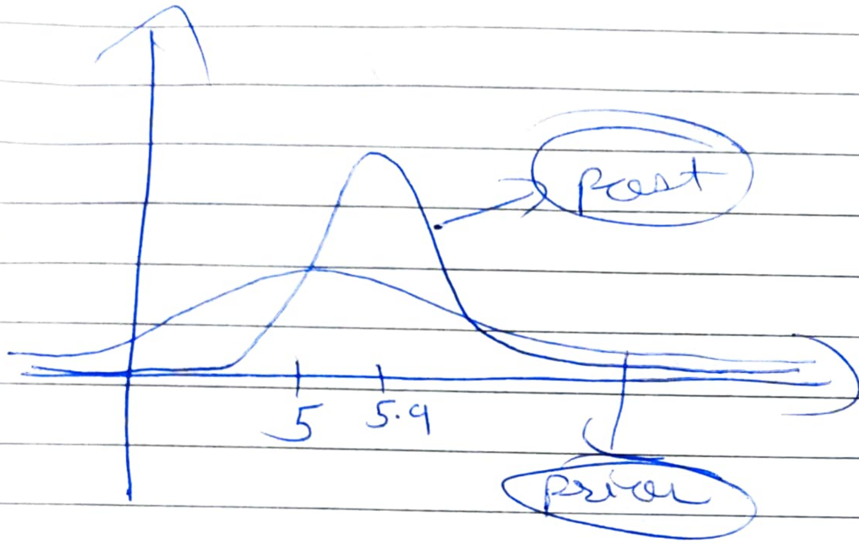
$$\sigma^2\ post = \frac{1}{a+b} = \frac{9}{10} = 0.9$$

So, post $(\theta|x) \sim N(5.9, 0.9)$

Teacher's Signature

Notes

Date / /
Page No.



The Data has Changed our believes about the Distribution of x, Making it more -ve Skewed and less flatter (Higher Kurtosis {by karl pearson coeffi}), reflecting the effect of our Data.

(e) as Content with Data (finite)

$b = \frac{n}{\sigma^2}$

as $\lim_{n \to \infty} b \to \infty$

so, $\lim_{n \to \infty} \mu_{post} = \bar{x}$, $\lim_{n \to \infty} \sigma^2_{post} \to 0$

Teacher's Signature

Notes

So as Data Increases ~~...~~ Variance of posterior tends towards Zero and μ_{post} converges to Sample Mean i.e the Data allows us to predict the original θ with more and more accuracy with more Data.

By Sen

$$\lim_{n \to \infty} \theta \rightarrow \bar{x}$$

So, sending same signal multiple times allows us to predict the original signal with higher precision

(d) $IQ_{General} \sim N(100, 15^2)$

$IQ_{test} \mid IQ_{General} \sim N(IQ_{General}, 10^2)$

So, $(IQ_{test} - IQ_{General}) \sim N(0, 10^2)$

Let $IQ_{General}$ be X. and IQ_{test} be Y.

then $X \sim N(100, 15^2)$
$Y \mid X \sim N(X, 10^2)$

OR

$(Y - X) \sim N(0, 10^2)$

So, we need to find $\cancel{E[X|Y]}$

$E[X \mid Y = y] = ?$

Using Normal-Normal updating,

$$\mu_{post} = \frac{\dfrac{\mu_{prior}}{\sigma^2_{prior}} + \dfrac{n\bar{X}}{\sigma^2_{lik}}}{\dfrac{1}{\sigma^2_p} + \dfrac{n}{\sigma^2_{lik}}}$$

$$= \frac{\dfrac{100}{15^2} + \dfrac{80}{10^2}}{\dfrac{1}{15^2} + \dfrac{1}{10^2}} = 88.031496063$$

Notes

(ii) $$\mu_{post} = \frac{\frac{100}{152} + \frac{150}{102}}{\frac{1}{152} + \frac{1}{102}} = 129.92126$$

Feels like a Bayesian trap :)

(Q.3) We will ~~take~~ Maximize the partial derivatives of L.L. for both parameters.

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \quad \left(f(x_i;\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right)$$

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

So, $l(\mu, \sigma^2) = -n \log \sqrt{2\pi}\sigma - \sum \frac{(x_i-\mu)^2}{2\sigma^2}$

$$\frac{\partial l}{\partial \mu} = -\sum \frac{-2(x_i-\mu)}{2\sigma^2} = 0 \Rightarrow \sum x_i = n\mu$$

$\Rightarrow \mu_{MLE} = \bar{X} \rightarrow$ Sample Mean

Teacher's Signature

Notes

Let $\theta = \sigma^2$

So, $\dfrac{\partial l}{\partial \theta} = \dfrac{\partial\left(-n\log\sqrt{2\pi\theta}\right)}{\partial \theta} - \dfrac{\partial\left(\sum\frac{(x_i-\mu)^2}{2\theta}\right)}{\partial \theta}$

$$\Rightarrow \frac{\partial l}{\partial \theta} = \frac{-n}{\sqrt{2\pi\theta}} \times \frac{\sqrt{2\pi}}{2\theta} + \frac{1}{2\theta^2}\sum(x_i-\mu)^2 = 0$$

$$\Rightarrow \frac{\partial l}{\partial \theta} = \frac{-n}{2\theta} + \frac{1}{2\theta^2}\sum(x_i-\mu)^2 = 0$$

$$\Rightarrow \theta = \frac{\sum(x_i-\mu)^2}{n}$$

$$\Rightarrow \theta_{MLE} \text{ or } \sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

and

$$\mu_{MLE} = \bar{x}$$

So, we can take sample mean as mean and σ^2_{MLE} as given in the formula.

Q.4) outputs $\rightarrow \{1, -1\}$

$$\hat{\theta}_{MAP} = \arg\max_{\theta}\left(\log P(y|\theta) + \log P(\theta)\right)$$

parameters of logistic regression

$H_\theta(x) = \dfrac{1}{1+e^{-t}}$ where $t = ax+b$

$P(y=1|x) = H(x)$

$P(y=-1|x) = 1-H(x)$

So, the Parameters are a, b.

$$P(y|x;\theta) = h(x)^y \, (1-h(x))^{1-y}$$

where $h(x) = \dfrac{1}{1+e^{-(ax+b)}}$

$$L(\theta) = \prod P(y|x;\theta)$$

Prior $\Rightarrow$ $P(a,b) = \dfrac{1}{2\pi} |\Sigma|^{-1/2} \exp\left\{-\dfrac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right\}$

Let $\theta = a, b$

Posterior $\Rightarrow$ $P(\theta|x) \propto P(y|x;\theta)\, P(\theta)$

Now MAP is the $\hat{\theta}$ such that

$$\underset{\theta}{\arg\max}\left(\log P(y|x;\theta) + \log P(\theta)\right).$$

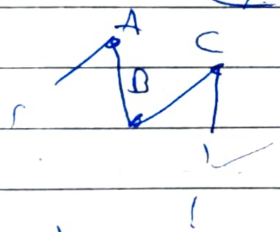
We can use Gradient Descent or Fisher Method to calculate this.

Q.5)

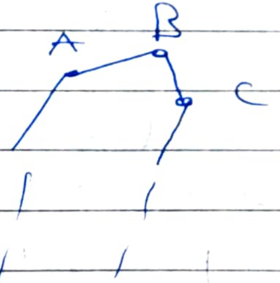
(a) A constant function gives only 1 and same classification every time. So, $VC = 0$.

(b) d+2 polygon

2 cases



OR



We can't get B alone by using a d dimension Hyperplane in this case

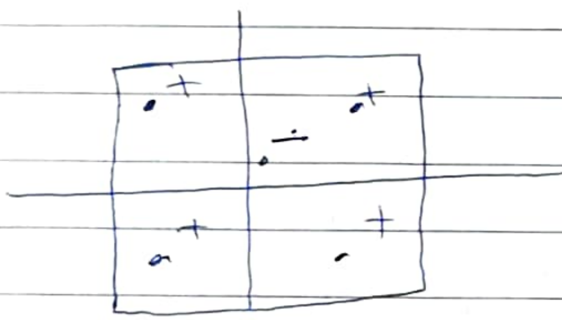
can't have A and C be on one side of Hyperplane and B on another side.

Notes

But, We can cut d+1 Polygon vertices in any 2 Sets with d dimension Hyperplane,
So, VC dimension = d+1.

(c)

~~In a 5~~



In a 5 Set Case, It is Not possible to have All 4 outer Most points in 1 Set and Inner point in other Set of Classification

So, VC = 4

(d) In Case of 3 point x_1, x_2, x_3 where $x_1 \le x_2 \le x_3$ If x_1 and x_3 are Inside Interval and x_2 has to be Inside too. So, (+ − +) Case is not possible
VC = 2