Q.1

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$\cancel{P(A_c B/T)}$

$$P(A_c B/T) = \frac{P(T/A_c B)\cdot P(A_c B)}{P(A/A_c B)P(A_c B) + P(T/A_s)P(A_s)}$$

$$= \frac{1}{1+\frac{P(T/B)\cdot P(A_c B)}{P(T/A)\cdot P(A_c B)}}$$

$$P(A_c B) = \frac{1}{11} \quad \text{— (1)}$$

$$= \frac{1}{1+\left(\frac{\lambda_B}{\lambda_A}\right)^{\sum_{i=1}^{5} k_i} e^{-(\lambda_B-\lambda_A)\times 20}}$$

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

$$\boxed{Ans = 0.96}$$

Q-2 (a) We know,

$$\text{Posterior} = \frac{\text{Prior Likelihood}}{\text{Marginal Likelihood}} \quad \text{—} (1)$$

$$P(\theta|X) \propto P(X|\theta) \cdot P(\theta) \quad \text{—} (2)$$

In given situation for us,

Prior $\sim \mathcal{N}(5, 9)$ & Likelihood $\sim \mathcal{N}(\theta, 4)$.

Also, $X = 6$

$$\therefore P(\theta|6) \propto \frac{e^{-\frac{(6-\theta)^2}{8}}}{\sqrt{8\pi}} \cdot \frac{e^{-\frac{(5-\theta)^2}{18}}}{\sqrt{18\pi}}$$

$$\propto \frac{e^{-\frac{13\theta^2 - 148\theta}{72}}}{\sqrt{8\pi} \cdot \sqrt{18\pi}}$$

Now, $\int_{-\infty}^{\infty} P(\theta|6) = 1$

$$\therefore = \frac{e^{-\frac{\left(\theta - \frac{74}{13}\right)^2}{2 \cdot \frac{36}{13}}}}{\sqrt{2\pi \cdot \frac{36}{13}}}$$

So, Posterior $\sim \mathcal{N}(74/13, 36/13)$

Q.2(b) $$\mu_{posterior} = \frac{\frac{\mu_{prior}}{\sigma^2_{prior}} + \frac{n\bar{X}}{\sigma^2_{likelihood}}}{\frac{1}{\sigma^2_{prior}} + \frac{n}{\sigma^2_{likelihood}}} \quad -① $$

from Bayesian Normal Updating

Now, Just putting values,

$$= \frac{\frac{5}{9} + \frac{4\times6}{4}}{\frac{1}{9} + \frac{4}{4}} \Rightarrow \boxed{\mu_{posterior} = 5.9}$$

$$\sigma^2_{posterior} = \frac{1}{\frac{1}{\sigma^2_{prior}} + \frac{n}{\sigma^2_{likelihood}}}$$

$$= \frac{1}{\frac{1}{9} + \frac{4}{4}}$$

$$\boxed{\sigma^2_{posterior} = 0.9}$$

Q.2(d)(i) For Randall Vard Test

R-Real
T-Test

Posterior = $\frac{\text{Prior Likelihood}}{\text{Marginal Likelihood}}$

$$P(IQ_R|IQ_T) \propto P(IQ_T|IQ_R) \cdot P(IQ_R)$$

Prior $\sim \mathcal{N}(100, 15^2)$, Likelihood $\sim \mathcal{N}(\overset{\nearrow 80}{IQ_T}, 10^2)$

$$\mu_{posterior} = \frac{\frac{\mu_{prior}}{\sigma^2_{prior}} + \frac{n\overline{X}}{\sigma^2_{likelihood}}}{\frac{1}{\sigma^2_{prior}} + \frac{n}{\sigma^2_{likelihood}}}$$

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

$$= 88.031$$

$\therefore$ Randall Ward's expected IQ is 88.031.

Q.2.(d)(ii)

Prior $\sim \mathcal{N}(100, 15^2)$, Likelihood $\sim \mathcal{N}(IQ, 10^2)$
$IQ \to 150$

$$\mu_{posterior} = \frac{\frac{100}{15^2} + \frac{1 \times 150}{10^2}}{\frac{1}{15^2} + \frac{1}{10^2}}$$

$$= 129.921$$

$\therefore$ Mary J. Taft's expected IQ is ~ 129.92

Q.3: The likelihood func^n for a sample of size 'n' is:

$$L(X|(\mu,\sigma^2)) = P(X|(\mu,\sigma^2)) \quad \text{--- (1)}$$

$$= \prod_{i=1}^{n} \frac{e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

$$= \frac{e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}}{\left(\sqrt{2\pi\sigma^2}\right)^n}$$

Now, For the log likelihood,

$$l(X|(\mu,\sigma^2)) = \ln(L(X|(\mu,\sigma^2)))$$

$$= -\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} - \frac{n}{2}\ln(2\pi\sigma^2) \quad \text{--- (2)}$$

For estimating parameters using MLE, the following condit must be satisfied

$$\frac{\partial l}{\partial \mu} = 0, \frac{\partial l}{\partial \sigma} = 0, \frac{\partial^2 l}{\partial \mu^2} < 0, \frac{\partial^2 l}{\partial \sigma^2} \leq 0 \quad \text{--- (3)}$$

Thus, $\boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}$ & $\boxed{\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n}}}$

Scanned with OKEN Scanner

Q.4b–

$$\frac{\partial F}{\partial \theta} = \sum_{i=1}^{m} \left(y_i - h(\theta, x^2)\right) x^i - \sum_{i=1}^{m} \Sigma^{-1}(\theta - \mu)$$

$\hookrightarrow$ (1)

argmax θ :

$\hookrightarrow \theta_i = \theta + \alpha \frac{\partial F}{\partial \theta}$

$$\theta_i = \theta + \alpha \left[(1) \right]$$

$$P(y^2 / \theta, x^2) = \left(h(\theta, x^i)\right)^{y^i} \times \left(1 - h(\theta, x^1)\right)^{1-y^i}$$

$$P(\theta) = \frac{1}{\sqrt{2\pi} \, |\Sigma|^{d/2}} e^{-\frac{1}{2}(x-\mu^T)\Sigma^{-1}(x-\mu)}$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\text{argmax}} \log P(y^i / \theta, x^i) \, p(\theta)$$

$$h(\theta, x^i) = \frac{1}{1 + e^{-\theta^T x^i}}$$

$$f = \sum_{i=1}^{m} y^i \log(h(\theta, x^i)) + (1 - y^2) \log(1 - h(\theta, x^i)) + \log\left(\frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{1/2}}\right) - \frac{1}{2}(\theta - \mu)^T \Sigma^{-1} (\theta - \mu)$$

$$\frac{\partial f}{\partial \theta} = \sum_{i=1}^{m} (y_i - h(\theta x^i)) x^i - m \, \Sigma^{-1} (\theta - \mu)$$

Using Gradient Descent $\quad \theta_i = \theta + \alpha \frac{\partial F}{\partial \theta}$

$$1D : \theta = \begin{bmatrix} \theta_i \\ \theta_m \end{bmatrix}$$

$$\frac{\partial F}{\partial \theta_0} = \sum_{i=1}^{m} (y^i - h(\theta, x^i)) \cancel{x_0^i} \rightarrow 1$$

$$- m \sum \sigma_{b_1 b_0} (\theta_1 - \mu_1)$$

$$- \sigma^2_{b_0} (\theta_0 - \mu_0)$$

$\sigma_{b_1 b_0} = 0$ (Assuming b_1, b_0 to be independent)

$$\therefore \frac{\partial F}{\partial \theta_{b_1}} = \sum_{i=1}^{m} (y^i - h(\theta, x^i)) x^i_{b_1} - \sigma^2_{b_0} (\theta_0 - \mu)$$

Q.5 (a) Constant function — we can classify every point singly.

VC dimension = 0.

(b) $$f^n_{x \in R^d} = \omega^T x$$

$$x^i_k = 1_{i=k}$$

$$S = \{x^i\}, \ \forall i \in \{1, \ldots, d\}$$

Let the labelling be $\{y^i\}$ $\forall$ $i \in \{1, \ldots, d\}$

$$f(x) = \text{sign}\left\{\sum_{i=1}^{d} y^i x^{iT} x\right\}$$

$f(x^i) = y^i$ & we can specify values of y^i for all different classification.

VC dim $\geq d$

Let's consider $(d+1)$ points

We know, $x^{d+1} = \sum_{i=1}^{d} a_i x_i$ (assuming x^i LI)

$$\omega^T x^{d+1} = \sum_{i=1}^{d} a_i \omega_i^T x^i$$

$\therefore$ This classification of x^{d+1} is fixed when all the others have been classified.

No 'w' can shatter (d+1) points

$\therefore \boxed{\text{VC dim} = d}$

VC dim of affine fun = d+1 (bcz it is linear in (d+1) dimens")

(c) Let us consider this classification
$S = \{x_1, x_2, \text{———} x_5\}$

$\underbrace{X_{max}, X_{min}, Y_{max}, Y_{min}}_{2}$

Points having these values are classified as '+' and other as −

Max. no. of pts = 4 but an axis aligned rectangle containing these 4 will always have the +ve hsh

$\therefore \boxed{\text{VC dim} = 5}$

(d) Intervals 1d

No interval (Not disjoint can do these classific")

$\therefore \boxed{\text{VC dim} = 2}$