HOMEWORK ASSIGNMENT - 1

* Name : DAKSH BHUVA
* CWID : 10475468

* Problem 1. Probability

Major	Session 1	Session 2	Session 3
CS	6	10	6
STAT	8	10	6
MG	6	0	8

⇒ Given $p(S1) = 0.2$, $p(S2) = 0.2$, $p(S3) = 0.6$

i) Probability that selected student has Major in CS:

$\therefore p(CS) = p(CS, S1) + p(CS, S2) + p(CS, S3)$

Marginal Distribution

we know, $p(CS, S1) = p(CS \mid S1) \cdot p(S1)$

Conditional Distribution

$$\therefore P(CS) = p(CS \mid S1) \cdot p(S1) + p(CS \mid S2) \cdot p(S2) + p(CS \mid S3) \cdot p(S3)$$

$$= \frac{6}{20} \cdot \frac{2}{10} + \frac{10}{20} \cdot \frac{2}{10} + \frac{6}{20} \cdot \frac{6}{10}$$

$$= \frac{6}{100} + \frac{10}{100} + \frac{18}{100} = \boxed{\frac{34}{100}}$$

$\therefore p(CS) = \boxed{0.34}$

(2) Probability that student is from session 3, given student is from STAT:

$\therefore$ $$p(STAT \mid S3) = \frac{p(S3 \mid STAT) \cdot p(Stat)}{p(S3)}$$

Because $$p(S3 \mid STAT) = \frac{p(S3 \cap STAT)}{p(STAT)}$$

$\therefore$ $p(S3 \cap STAT) = p(S3 \mid STAT) \cdot p(STAT)$

$\therefore$ $$p(S3 \mid STAT) = \frac{p(STAT \mid S3) \cdot p(S3)}{p(STAT)} \quad \text{.....(i)}$$

→ Now $p(STAT \mid S3) = \frac{6}{20}$

→ Also, $p(STAT) = p(STAT, S1) + p(STAT, S2) + p(STAT, S3)$

$= p(STAT \mid S1) \cdot p(S1) + p(STAT \mid S2) \cdot p(S2) + p(STAT \mid S3) \cdot p(S3)$

$= \frac{8}{20} \cdot \frac{2}{10} + \frac{10}{20} \cdot \frac{2}{10} + \frac{6}{20} \cdot \frac{6}{10}$

$= \frac{36}{100}$

→ Now, putting above values in equation(i).

$$\therefore p(S3 \mid STAT) = \frac{p(STAT \mid S3) \cdot p(S3)}{p(STAT)}$$

$$= \frac{\frac{6}{20} \cdot \frac{6}{10}}{\frac{36}{100}} = \boxed{\frac{1}{2}}$$

$$\therefore p(S3 \mid STAT) = \boxed{0.5}$$

* Problem 2. Maximum Likelihood Estimation (MLE) :-

→ Given weights :- 112, 120, 131, 126, 145, 158, 157, 136, 148, 176

* ~~[illegible]~~ ~~[illegible]~~ ~~[illegible]~~

(1) Probability of dataset given the two parameters.

$$\therefore P(x \mid \mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n \mid \mu, \sigma^2)$$

$$; \text{ where } \mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$\therefore$ Likelihood function :

$$P(x \mid \mu, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^5} \exp\left(\frac{-1}{2}\sum_{n=1}^{N}\left(\frac{x_N-\mu}{\sigma}\right)^2\right)$$

(2) → Applying log on both the sides:-

$$\log\left(P(x \mid \mu, \sigma^2)\right) = -5\log(2\pi) - 5\log(\sigma^2) - \frac{1}{2}\sum_{n=1}^{N}\left(\frac{x_N-\mu}{\sigma^2}\right)^2$$

classmate

Date ____
Page ____

→ Now, derivating log $(p(x|\mu, \sigma^2))$ w.r.t μ and equating it to 0,

$$\therefore \frac{\partial \log(p(x|\mu,\sigma^2))}{\partial \mu} = \frac{\partial\left[-5\log(2\pi) - 5\log(\sigma^2) - \frac{1}{2}\sum_{n=1}^{N}\left(\frac{x_n-\mu}{\sigma}\right)^2\right]}{\partial \mu}$$

$$\therefore \Rightarrow \frac{\partial\left(-\frac{1}{2}\sum_{n=1}^{N}\left(\frac{x_n-\mu}{\sigma}\right)^2\right)}{\partial \mu} = 0$$

$$\therefore \Rightarrow -\frac{1}{2}\sum_{n=1}^{N} \frac{2(x_n-\mu)(-1)}{\sigma^2} = 0$$

$$\therefore \Rightarrow \sum_{n=1}^{N}(x_n - \mu) = 0$$

$$\therefore \Rightarrow N\mu = \sum_{n=1}^{N} x_n$$

$$\therefore \mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

⇒ Maximum likelihood estimation, (MLE) of μ:

$$(MLE)_\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

classmate

Date

Page

$$MLE = \frac{1}{10}\left(112 + 120 + 131 + 126 + 145 + 158 + 157 + 136 + 148 + 176\right)$$

$$= \frac{1}{10}(1409)$$

$$MLE = \boxed{140.9} \text{ pounds.}$$

→ Now, derivating $\log(P(x|\mu,\sigma^2))$ w.r.t σ^2 and equating it to 0,

$$\therefore \frac{\partial}{\partial\sigma^2}\left(\log(P(x|\mu,\sigma^2))\right)$$

$$= \frac{\partial}{\partial\sigma^2}\left(-5\log(\sigma^2) - \frac{1}{2}\sum_{n=1}^{N}\left(\frac{x_n-\mu}{\sigma}\right)^2\right)$$

$$\therefore \Rightarrow -\frac{5}{\sigma^2} - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2\left(\frac{-1}{(\sigma^2)^2}\right) = 0$$

$$\Rightarrow \frac{1}{2}\sum_{n=1}^{N}\frac{(x_n-\mu)^2}{(\sigma^2)^2} - \frac{5}{\sigma^2} = 0$$

$$\therefore \Rightarrow \frac{5}{\sigma^2} = \frac{1}{2}\sum_{n=1}^{N}\frac{(x_n-\mu)^2}{(\sigma^2)^2}$$

$$\therefore \Rightarrow \sigma^2 = \frac{1}{10}\sum_{n=1}^{N}(x_n-\mu)^2$$

classmate
Date
Page

∴ Hence, MLE of σ^2:

$$(MLE)_{\sigma^2} = \frac{1}{10} \sum_{n=1}^{N} (x_n - \mu)^2$$

$$\therefore \quad = \frac{1}{10} \sum_{n=1}^{N} (x_n - 140.9)^2$$

$$= \frac{1}{10} \Big[(-28.9)^2 + (-20.9)^2 + (-9.9)^2 + (-14.9)^2 + (4.5)^2 + (17.1)^2 + (16.1)^2 + (-4.9)^2 + (7.1)^2 + (35.1)^2 \Big]$$

$$= \frac{1}{10} [3466.9]$$

$$(MLE)_{\sigma^2} = \boxed{346.69}$$

Date
Page

* Problem 3. ~~ab~~ MLE

X	1	2	3	4
$P(x)$	$\frac{2q}{3}$	$\frac{q}{3}$	$\frac{2(1-q)}{3}$	$\frac{(1-q)}{3}$

→ obs : 4, 1, 3, 2, 4, 3, 2, 1, 3, 2

(1) Likelihood function :-

$$\therefore P(q) = \prod_{n=1}^{N} P(\text{obs}_n)$$

$$= P(4) \cdot P(1) \cdot P(3) \cdot P(2) \cdot P(4) \cdot P(3) \cdot P(2) \cdot P(1) \cdot P(3) \cdot P(2)$$

$$= \left(\frac{2a}{3}\right)^2 \left(\frac{q}{3}\right)^3 \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{1-q}{3}\right)^2$$

[From the above table]

⇒ Applying log on both sides,

$$\therefore \log(P(q)) = 2\log\left(\frac{2q}{3}\right) + 3\log\left(\frac{q}{3}\right) + 3\log\left(\frac{2(1-q)}{3}\right) + 2\log\frac{(1-q)}{3}$$

$$\therefore \log(P(q)) = 7\log 2 - 10\log 3 + 5\log q + 5\log(1-q)$$

→ Now, derivating w.r.t. q, equating to 0:

$$\therefore \frac{\partial}{\partial q}\left(\log\left(P(q)\right)\right) = \frac{\partial}{\partial q}\left(5\left(\log q + \log(1-q)\right)\right)$$

$$\therefore \Rightarrow \frac{5}{q} + \frac{5}{1-q}(-1) = 0$$

$$\therefore \Rightarrow \frac{5}{q} - \frac{5}{1-q} = 0$$

$$\therefore \Rightarrow \frac{5}{q} = \frac{5}{1-q}$$

$$\therefore \boxed{q = \frac{1}{2} = 0.5}$$

* Problem 4. MAP estimation

$\therefore p(y|x, w, \beta) = \mathcal{N}(y | f(x, w), \beta^{-1})$

$\therefore p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$

$\therefore p(y|x, w, \beta) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp\left(\frac{-1}{2} \frac{(y_n - f(x_n, w))^2}{\beta^{-1}}\right)$

$\Rightarrow \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

$\therefore p(y|x, w, \beta) = \prod_{n=1}^{N} \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp\left(\frac{-\beta}{2}(y_n - f(x_n, w))^2\right)$

$\therefore p(w|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$

→ using "bayes theorem", we know,

posterior $\propto$ likelihood $\times$ prior

$\therefore p(w|x, y, \alpha, \beta) = (k) \cdot p(y|x, w, \beta) \cdot p(w|\alpha)$

classmate

Date

Page

→ Applying log on both sides,

(i) $\log(p(w | x, y, \alpha, \beta)) = \log k + \log(p(y | x, w, \beta))$
$+ \log(p(w | \alpha))$

→ Now, $p(y | x, w, \beta) = \prod_{n=1}^{N} \frac{\sqrt{\beta}}{\sqrt{2\pi}} \exp\left(\frac{-\beta}{2}(y_n - f(x_n; w))^2\right)$

$\therefore \log(p(y | x, w, \beta)) = \log\left(\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \exp\left(\frac{-\beta}{2} \sum_{n=1}^{N} (y_n - f(x_n, w))^2\right)\right)$

$= \frac{N}{2}\log(\beta) - \frac{N}{2}\log(2\pi) - \frac{\beta}{2}\sum_{n=1}^{N} (y_n - f(x_n, w))^2$

$\therefore p(w | \alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{m+1}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$

$\therefore \log(p(w | \alpha)) = \log\left(\left(\frac{\alpha}{2\pi}\right)^{\frac{m+1}{2}} \exp\left(\frac{-\alpha}{2} w^T w\right)\right)$

$= \frac{(m+1)}{2}\log\alpha - \frac{(m+1)}{2}\log(2\pi)$
$- \frac{\alpha}{2} w^T w$

→ Substituting above values in (i),

$$\therefore = \log k + \frac{N}{2}\log\beta - \frac{N}{2}\log(2\pi) - \frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2$$

$$+ \frac{M+1}{2}\log\alpha - \frac{(M+1)}{2}\log(2\pi) - \frac{\alpha}{2} w^T w$$

$$= \text{Constant} - \left(\frac{\beta}{2}\sum_{n=1}^{N}(y_n - f(x_n, w))^2 + \frac{\alpha}{2} w^T w\right)$$

→ Taking negative log, thus
-log (posterior) needs to be minimized.

$$\therefore -\log(\text{posterior}) = \frac{\beta}{2}\sum_{n=1}^{N}(f(x_n, w) - y_n)^2 + \frac{\alpha}{2} w^T w - \text{Constant}$$

→ This would give us maximum posterior

→ This is regularized sum-of-squares error function

∴ Hence proved that Maximum posterior (MAP) is equivalent to minimizing the regularized sum-of-squares error function.