Assignment - 1.

Q1(a) To Prove :- The least square fit line always passes through the point $(\bar{x}, \bar{y})$, where $\bar{x}$ & $\bar{y}$ represent mean of X & Y.

To show $\bar{y}$ & $\bar{x}$ lie on the line we need to prove $\bar{y} = w_1 \bar{x} + b$ where w_1 & b are parameters obtained from regression.
So, let y^i represent value of y for i^{th} sample & same follows for x^i

Let the linear regression model be
$$\hat{y} = w_1 x + b$$

Now $y^{(i)} = \hat{y}^{(i)} + \epsilon_i$ (True value = Predicted value + Error).

$$y^{(i)} = w_1 x^{(i)} + \epsilon_i + b$$

$$y^{(1)} = w_1 x^1 + b + \epsilon_1$$
$$y^{(2)} = w_1 x^{(2)} + b + \epsilon_2$$
$$\vdots$$
$$y^{(n)} = w_1 x^{(n)} + b + \epsilon_n$$

Summing above
$$\sum_{i=1}^{n} y^{(i)} = w_1 \sum_{i=1}^{n} x^{(i)} + bn + \sum_{i=1}^{n} \epsilon_i$$

dividing by n
$$\frac{1}{n} \sum_{i=1}^{n} y^{(i)} = \frac{w_1}{n} \sum_{i=1}^{n} x^{(i)} + \frac{bn}{n} + \frac{1}{n} \sum_{i=1}^{n} \epsilon_i$$

So, ~~sum~~ to show that $\sum_{i=1}^{n} \epsilon_i = 0$.

Sum of mean square erros = $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$$MSE = \sum_{i=1}^{n} (y_i - w_1 x_i - b)^2$$

For this to be min $\frac{\partial MSE}{\partial b} = 0$

$\Rightarrow \sum_{i=1}^{n} 2(y_i - w_1 x_i - b) \cdot -1 = 0$.

$\Rightarrow$ Now $\epsilon_i = y_i - w_1 x_i - b$

so $-2\sum_{i=1}^{n} \epsilon_i = 0 \Rightarrow \sum_{i=1}^{n} \epsilon_i = 0$

So, $\bar{y} = w_1 \bar{x} + b + \bar{\epsilon}$

$\bar{y} = w_1 \bar{x} + b$ (as $\bar{\epsilon} = 0$)

Hence $(\bar{y}, \bar{x})$ lie on the least square fit line.

classmate

Date
Page

Q1 (b). Here let

$\rho_{x,y}$ denote correlation coefficient between x & y

Now, given

$\rho_{x,z} > 0$

$\rho_{y,z} > 0$

Now we claim $\rho_{x,y}$ need not be greater than 0 always

So, using partial correlation,

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{xz} \cdot \rho_{zy}}{\sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{zy}}}$$

$$\Rightarrow \rho_{xy} = \rho_{xz} \cdot \rho_{yz} + \rho_{xy.z}\sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}$$

Now by condition of partial coeff. $-1 \leq \rho_{xy.z} \leq 1$.

So

$$\rho_{xy} = \rho_{xz} \cdot \rho_{yz} \pm \cancel{\rho_{xy.z}} \sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}$$

So the upper & lower band of ρ_{xy} are

$\rho_{xz} \cdot \rho_{yz} + \sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}$ & $\rho_{xz} \cdot \rho_{yz} - \sqrt{1-\rho^2_{xz}}\sqrt{1-\rho^2_{yz}}$

respectively.

Now assuming $\rho_{xz} = 0.75$ & $\rho_{yz} = 0.8$ which indicates strong correlation b/w them

$$\rho_{xy} \in \left[(0.75)(0.8) - \sqrt{1-0.75^2}\sqrt{1-0.8^2},\ (0.75)(0.8) + \sqrt{1-0.75^2}\sqrt{1-0.8^2} \right]$$

$$\rho_{xy} \in \left[0.6 - \sqrt{0.4375}\sqrt{0.36},\ 0.6 + \sqrt{0.4375}\sqrt{0.36} \right]$$

$$\rho_{x,y} \in [0.6 - 0.39,\ 0.6 + 0.39]$$

$$\rho_{x,y} \in [0.21, 0.99]$$

So $\rho_{x,y}$ can be between 0.21 to 0.3 as well. which indicates a very weak correlation b/w x, y.

Hence a strong correlation is not guaranteed always

Hence the proof.

Q1(c) Given:- $X_1, X_2, X_3 \ldots X_n$ a sequence of iid random variables.

$E(X_1) = E(X_2) = \ldots E(X_n) = \mu$

To Prove :- For $n \to \infty$ $\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots X_n}{n} \to \mu$

ie

$\cancel{P\left(\left[\left|\frac{X_1+X_2+\ldots X_n}{n}\right| - \mu\right]\right)}$;

ie let $S_n = \frac{X_1 + X_2 + \ldots X_n}{1}$

$\lim_{n\to\infty} P\left(\left(\left|\frac{S_n}{n} - \mu\right|\right) \geq \epsilon\right) \longrightarrow 0$ for any $\epsilon > 0$

Proof

$S_n = X_1 + X_2 + \ldots + X_n$

$\bar{S_n} = \frac{S_n}{n}$

$\bar{S_n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$

$\bar{S_n} = \frac{\mu n}{n}$

$\bar{S_n} = \mu$

classmate

Date
Page

Now, By Chebyshev's inequality,

$$P\left[|\bar{S}_n - \mu| \geq \epsilon \right] \leq \frac{Var(\bar{S}_n)}{\epsilon^2}$$

Now

$$Var(\bar{S}_n) = Var\left(\frac{X_1 + X_2 + \dots + X_n}{n} \right)$$

$$= \frac{1}{n^2} Var(X_1 + X_2 + \dots + X_n)$$

$$= \frac{1}{n^2} [Var(X_1) + \dots Var(X_n)]$$ (as X_i are iid)

$$= \frac{n\sigma^2}{n^2}$$ (as $Var(X_1) = Var(X_2) = \dots Var(X_n) = \sigma^2$ as they are iid).

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

So, $$P\left[|\bar{S}_n - \mu| \geq \epsilon \right] \leq \frac{\sigma^2}{n\epsilon^2}$$

So $$\lim_{n \to \infty} P\left[|\bar{S}_n - \mu| \geq \epsilon \right] \leq \lim_{n \to \infty} \frac{\sigma^2}{n\epsilon^2}$$

$$\leq 0$$ (as $\lim_{n \to \infty} \frac{1}{n} = 0$)

Hence Proved.

Classmate

Date
Page

Q(a) <u>To Derive :-</u> Maximum A Posteriori (MAP) sol^n for linear regression

Now, $y = w^T x + \epsilon$

here

$\epsilon \sim \text{Normal}(0, \sigma^2)$

Hence $y|x \sim \text{Normal}(w^T x, \sigma^2)$

$$\Rightarrow P(y_i | x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i^T w - y_i)^2}{2\sigma^2}}$$

Now,

$$w = \underset{w}{\operatorname{argmax}} P(w | y_1, x_1, y_2, x_2 \ldots y_n, x_n)$$

$$w = \underset{w}{\operatorname{argmax}} \frac{P(y_1, x_1, \ldots y_n, x_n | w) P(w)}{P(y_1, x_1, \ldots y_n, x_n)} \quad \text{(Bayes Rule)}$$

$$= \underset{w}{\operatorname{argmax}} P(y_1, x_1, \ldots y_n, x_n | w) P(w).$$

Now we assume w to be $w \sim N(\vec{0}, \tau^2 I)$.

$$\text{So, } P(w) = \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2}\left(\frac{w-0}{\tau}\right)^2}$$

$$P(w) = \frac{1}{\sqrt{2\pi\tau^2}} \cdot e^{\frac{-w^2}{2\tau^2}}$$

Now $P(y_1, x_1 \cdots y_n, x_n | w) = \prod_{i=1}^{n} P(y_i, x_i | w)$

Now ~~$P(x, y | z) = P$~~

$P(y, x | z) = P(y | x, z) \, P(x | z)$

(as y_i depends on x_i & are iid)

So, $P(y_1, x_1 \cdots y_n, x_n | w) = \prod_{i=1}^{n} \left[P(y_i | x_i, w) \, P(x_i | w) \right]$

$$w = \underset{w}{\arg\max} \prod_{i=1}^{n} P[(y_i | x_i, w)] \, P(x_i | w) \cdot P(w)$$

$$= \underset{w}{\arg\max} \left[\prod_{i=1}^{n} P[y_i | x_i, w] \, P(x_i) \right] \cdot P(w).$$

(as x_i independant of w).

$$= \underset{w}{\arg\max} \sum_{i=1}^{n} \left[\log P(y_i | x_i, w) + \log P(w) \right] \rightarrow$$

(as $P(x_i)$ is constant)

$$= \underset{w}{\arg\max} \sum_{i=1}^{n} \left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) + \log\left(e^{\frac{-(x_i^T w - y_i)^2}{2\sigma^2}} \right) \right] + \left[\log\left(\frac{-w^2}{2\tau^2} \right) + \log\left(\frac{1}{\sqrt{2\pi\tau^2}} \right) \right]$$

$=$

$$= \underset{w}{\arg\min} \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \frac{1}{2\tau^2} w^T w$$

as $\log\left(\frac{1}{\sqrt{2\pi\tau^2}} \right)$ is constant so ignored.

$$\boxed{w = \underset{w}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \lambda (w^T w)}$$

$\left(\text{where } \lambda = \frac{\sigma^2}{n\tau^2} \right).$