Q1-a - 1

Q1

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in R$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = ?$$

$u = \frac{x}{\sqrt{2}}$ OR $x = \sqrt{2}u$ OR $dx = \sqrt{2}du$ — (1)

Limits of integration remain $-\infty$ to $+\infty$

Rewriting the integral in terms of u.

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-u^2} \sqrt{2} du = \sqrt{2} \times \int_{-\infty}^{+\infty} e^{-u} du \quad \text{— (2)}$$

Using the already known Gaussian integral

$$\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi} \quad \text{— (3)}$$

putting (1), (3) in (2)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \times \sqrt{2} \times \sqrt{\pi}$$

$$= \frac{\sqrt{2} \times \sqrt{\pi}}{\sqrt{2\pi}} = 1.$$

(b) X is Random variable with gaussian distribution

To prove $E[x]=0$

$$E[x] = \int_{-\infty}^{\infty} x \times f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x \times \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} x \times e^{-x^2/2}\, dx. \quad -② $$

Let $u = -x^2/2$ OR $du = -x\,dx$ OR $x\,dx = -du$

Limits

$x = -\infty \quad u = -\infty$

$x = \infty \quad u = -\infty$

So Re-writing ②

$$= \frac{1}{2\pi} \int_{-\infty}^{-\infty} e^{u}(-du)$$

Since limits are indentical, the integral eval to 0

So

$E[x]=0$

Q7-C.

$$VAR(X) = E[X^2] - (E[X])^2$$

We know $E[X]=0$ so

$$VAR(X) = E[X^2]$$

Expected val of $E[X^2]$ is.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx.$$

$$fx = \frac{1}{\sqrt{2\pi}} \times e^{-x^2/2}$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$E[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \times e^{-x^2/2} \quad — (1)$$

solving $\int_{-\infty}^{\infty} x^2 \times e^{-x^2/2}$ separately

Let $U = x \quad dV = xe^{-x/2}\,dx \quad — (2)$

If $dv = x \times e^{-x^2/2}\,dx$

$$V = \int xe^{-x^2/2}\,dx.$$

$$V = -e^{-x^2/2} \quad (3)$$

Also

$$du = dx. \quad (4)$$

Using integration by parts (using 2, 3, 4)

$$\int x^2 e^{-x^2/2} dx = uv - \int v\,du$$

Subst. $u = x$, $v = -e^{-x^2/2}$ $du = dx$

$$\int x^2 e^{-x^2/2} dz = x \times (-e^{-x^2/2}) - \int (-e^{-x^2/2})\,dx$$

• This becomes zero for $\pm\infty$

Using gaussian integral with scaling factor

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}.$$

similarly

$$\int x^2 e^{-x^2/2} dx = 0 - (-\sqrt{\pi} \times \sqrt{2})$$

$$= \sqrt{2\pi} \quad - 5$$

substitute 5 in 1

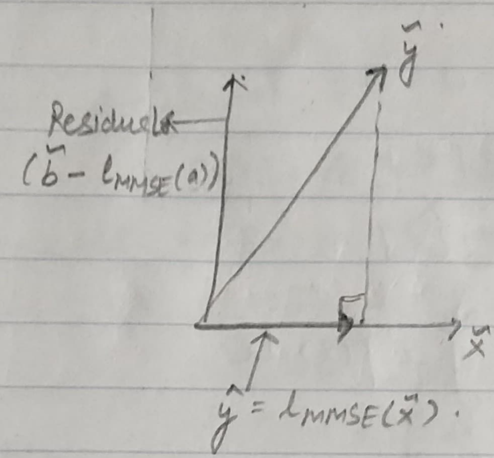
$$E[x^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \times e^{-x^2/2}$$

$$\boxed{E[x^2] = \frac{1}{\sqrt{2\pi}} \times \sqrt{2\pi} = 1}$$

Subst. values of $E[x]$ and $(E[x])^2$

$$\boxed{VAR(x) = E[x^2] - (E[x])^2 = 1 - 0 = 1}$$

Q2) The geometric intepretation of Regression is:-



When we first Regress $\vec{y}$ on $\vec{x}$, we find $\hat{y}$ which is in same sub-space as $\vec{x}$ (It minimizes the distance between $\vec{x}$ & $\vec{y}$),

Now once we have $\hat{y}$ (we will not be able to find a new better line) because Regressing $\hat{y}$ on x is like projecting $\hat{y}$ something that is already on the line back onto same line.

Q3-a For any $\alpha, \beta \in R$

$$cov[\beta\vec{a}+\alpha, \vec{b}] = \beta \, cov[\vec{a}, \vec{b}]$$

Covariance definition

$$cov[X,Y] = (E[X-E[X]])(E[Y-E[Y]]) \quad -①$$

substitute $X = \beta\vec{a}+\alpha \quad Y = \vec{b}$ in ①

$$cov[\beta\vec{a}+\alpha, \vec{b}] = E[(\beta\vec{a}+\alpha - E[\beta\vec{a}+\alpha])(\vec{b}-E[\vec{b}])] \quad -②$$

Simplify $E[\beta\vec{a}+\alpha] = \beta E[\vec{a}] + \alpha$

$$cov[\beta\vec{a}+\alpha, \vec{b}] = E[\beta(\vec{a}-E[\vec{a}])](\vec{b}-E[\vec{b}])$$

$$cov[\beta\vec{a}+\alpha, \vec{b}] = \beta E[\vec{a}-E[\vec{a}]](\vec{b}-E[\vec{b}])$$

$$cov[\beta\vec{a}+\alpha, \vec{b}] = \beta \, cov[\vec{a}, \vec{b}]$$

Hence proved.

Q3-b-1

To prove.

$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = VAR[\vec{a}] - VAR[\vec{b}]$

We know the defination of covariance

$$Cov[x,y] = E[x,y] - E[x]E[y] \quad — ①$$

Subst. $x = \vec{a}+\vec{b}$ $\quad y = \vec{a}-\vec{b}$ in ①

$$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = E[(\vec{a}+\vec{b})(\vec{a}-\vec{b})] - E[\vec{a}+\vec{b}]E[\vec{a}-\vec{b}]$$

Expanding $(\vec{a}+\vec{b})(\vec{a}-\vec{b})$.

$$(\vec{a}+\vec{b})(\vec{a}-\vec{b}) = \vec{a}^2 - \vec{b}^2$$

$$E[(\vec{a}+\vec{b})(\vec{a}-\vec{b})] = E[\vec{a}^2 - \vec{b}^2] = E[\vec{a}^2] - E[\vec{b}^2] \quad - ②$$

Expanding $E[\vec{a}+\vec{b}]E[\vec{a}-\vec{b}]$

$$E[\vec{a}+\vec{b}]E[\vec{a}-\vec{b}] = (E[\vec{a}]+E[\vec{b}])(E[\vec{a}]-E[\vec{b}])$$

$$E[\vec{a}+\vec{b}]E[\vec{a}-\vec{b}] = E[\vec{a}]E[\vec{a}] - E[\vec{b}]E[\vec{b}]$$

$$E[\vec{a}+\vec{b}]E[\vec{a}-\vec{b}] = (E[\vec{a}])^2 - (E[\vec{b}])^2 \quad — ③$$

Subst. ② & ③ in ①

$$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = (E[\vec{a}^2] - E[\vec{b}^2]) - ((E[\vec{a}])^2 - (E[\vec{b}])^2)$$

$$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = E[\vec{a}^2] - E[\vec{b}^2] - (E[\vec{a}])^2 - (E[\vec{b}])^2$$

$$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = [E[\vec{a}^2] - (E[\vec{a}])^2][E[\vec{b}^2] - (E[\vec{b}])^2]$$

$$Cov[\vec{a}+\vec{b}, \vec{a}-\vec{b}] = VAR(a)\ VAR(b).$$

Hence proved.

Q4 a-1

$$S(x_i) = \frac{x_i - m(x)}{\sqrt{V(x)}} \quad — ①$$

$m(x)$ is sample mean
$V(x)$ is sample variance.

$$S(y_i) = \frac{y_i - m(Y)}{\sqrt{V(Y)}}$$

To show $m(S_x) = 0$ $\quad m(S_y) = 0$

$$m(S_x) = \frac{1}{n} \sum_{i=1}^{n} S(x_i) \quad — ②$$

Subst. ② in ①

$$m(S_x) = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - m(x)}{\sqrt{V(x_i)}}$$

$$= \frac{1}{n} \times \frac{1}{\sqrt{V(x_i)}} \times \sum_{i=1}^{n} x_i - m(X)$$

$$= \frac{1}{n} \times \frac{1}{\sqrt{V(x_i)}} \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} m(X) \right)$$

$$= \frac{1}{n} \times \frac{1}{\sqrt{V(x_i)}} \left(\sum_{i=1}^{n} i - \underbrace{n \times m(X)} \right)$$

(subst. $m(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$= \frac{1}{n} \times \frac{1}{\sqrt{V(x_i)}} \left(\sum_{i=1}^{n} i - \sum_{i=1}^{n} i \right)$$

$$= 0$$

So $m(S_x) = 0$

Similarly $m(S_y) = 0$

To prove variance of standardized data is one.

$$V(S_x) = \frac{1}{n-1} \sum_{i=1}^{n} (S(x_i) - m(S_x))^2 \quad -①$$

We know $m(S_x) = 0$

subst. this in eqn ①

$$\frac{1}{n-1} \sum_{i=1}^{n} (S(x_i)^2)$$

Subst $S(x_i) = \frac{x_i - m(x_i)}{\sqrt{V(x)}}$.

$$V(S_x) = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - m(x)}{\sqrt{V(x)}} \right)^2$$

Factor out $= \frac{1}{\sqrt{V(x)}}$

$$V(S_x) = \frac{1}{n-1} \times \frac{1}{V(x)} \sum_{i=1}^{n} (x_i - m(x))^2$$

$$V(S_x) = \frac{1}{V(x)} \times \underbrace{\frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - m(x))^2}$$

$$V(S_x) = \frac{1}{V(x)} \times V(x)$$

$$V(S_x) = 1$$

similarly $V(S_y) = 1$

Sample covariance of standardized data equals correlation coefficient

$$c(s_x, s_y) = \frac{1}{n-1} \sum_{i=1}^{n} s(x_i) s(y_i) = \rho_{x,y}$$

Subst. $s(x_i) = \dfrac{x_i - m(x)}{\sqrt{V(x)}}$

$$s(y_i) = \frac{y_i - m(x)}{\sqrt{V(y)}}$$

$$c(s_x, s_y) = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(x_i - m(x))(y_i - m(y))}{\sqrt{V(x)}\sqrt{V(y)}}$$

Factor out $\dfrac{1}{\sqrt{V(x)}\sqrt{V(y)}}$.

$$c(s_x, s_y) = \frac{1}{\sqrt{V(x)}\sqrt{V(y)}} \times \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m(x))(y_i - m(y))$$

$$c(s_x, s_y) = \frac{c(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$$

We know sample correlation coefficient

$$\rho_{x,y} = \frac{c(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$$

$$c(s_x, s_y) = \rho_{x,y}$$

Q-4-b

4-b

To prove:

$$\frac{1}{n-1}\sum_{i=2}^{n} e_i^2 = (1 - P_{x,y}^2)\,V(Y).$$

Where

$e_i = y_i - l_{OLS}(x_i)$ Residual.

$P_{x,y}$ = sample correlation coefficient

$V(Y)$ is the sample variance of Y.

$$l_{OLS}(x_i) = \sqrt{V(\vec{y})} \times P_{\vec{x},\vec{y}} \left(\frac{\vec{x} - m(\vec{x})}{\sqrt{V(\vec{x})}} \right) + m(\vec{y})$$

$$l_{OLS}(x_i) = \sqrt{V(\vec{y})} \times P_{\vec{x},\vec{y}} \times S(\vec{x}) + m(\vec{y}) \quad \text{— (1)}$$

Now mean-squared Error.

$$m((l_{OLS}(y_i) - \vec{y}_i)^2). \quad \text{— (2)}$$

Replacing values of $l_{OLS}(x_i)$ from (1) into (2)

$$= m((\sqrt{V(\vec{y})} \times P_{\vec{x},\vec{y}} \times S(\vec{x}) + m(\vec{y}) - \vec{y})^2)$$

Converting $\vec{y}$ in $S(\vec{y})$.

$$= V(\vec{y})\, m((P_{\vec{x},\vec{y}} \times S(\vec{x}) - S(\vec{y}))^2).$$

$$= V(\vec{y})(P_{\vec{x},\vec{y}}^2\, m(S(\vec{x})^2) + m[S(\vec{y})^2] - 2P_{\vec{x},\vec{y}}\, m[S(\vec{x})S(\vec{y})]$$

$$= V(\vec{y})(P_{\vec{x},\vec{y}}^2 \times 1 + 1 - 2P_{\vec{x},\vec{y}}).$$

$$= V(\vec{y})(1 - P_{\vec{x},\vec{y}}^2).$$

Hence proved.

Q4-c.

To prove. $-1 \leq P_{x,y} \leq 1$.

$m((\tilde{y} - \ell_{OLS(x_i)}(\tilde{x}))^2) \geq 0$

From 4-b, subst. mean-squared error.

$v(\tilde{y})(1 - P^2_{\tilde{x},\tilde{y}}) \geq 0$

$1 - P^2_{\tilde{x}\tilde{y}} \geq 0$

$1 - P^2_{\tilde{x},\tilde{y}} \geq 0$

$P^2_{\tilde{x},\tilde{y}} \leq 1$

so -

$-1 \leq P^2_{\tilde{x},\tilde{y}} \leq 1$

Q4-d if $P_{x,y} = \pm 1$, then $y_i = \beta x_i + \alpha$ for some constant α, β

In a linear model, we know COD (R^2) is given by

$$R^2 = P^2_{x,y}$$

and

$$R^2 = 1 - \frac{SS_{Res}}{SS_{total}}.$$

Since CORR. coeff is ± 1.

SS_{total} can be 0 so $SS_{Residual} = 0$.

if

$$SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

if $SS_{res} = 0$ means there is NO residual

That is every data point (x_i, y_i), can be defined using a affine relation.

$$\hat{y} = \frac{\sqrt{VAR(\vec{y})} \times (\vec{x} - m(\vec{x}))}{\sqrt{VAR(\vec{x})}} + m(\vec{y})$$

Rearranging

$$\vec{y} = \vec{x} \times \underbrace{\frac{\sqrt{VAR(\vec{y})}}{\sqrt{VAR(x)}}}_{\beta} - \underbrace{\frac{m(\vec{x}) \times \sqrt{VAR(\vec{y})}}{\sqrt{VAR(\vec{x})}} + m(\vec{y})}_{\alpha}$$

Thus

$\vec{y} = \beta x + \alpha$. can be represented in this form.

Q6) $\hat{y}_i = x_i \hat{\beta}$ — ①

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i y_i\right) / \left(\sum_{i'=1}^{n} x_{i'}^2\right). \quad — ②$$

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'} \quad — ③$$

Subs. ② in ①

$$\hat{y}_i = x_i \left(\frac{\left(\sum_{i=1}^{n} x_i y_i\right)}{\left(\sum_{i'=1}^{n} x_{i'}^2\right)} \right)$$

$$\hat{y}_i = \left(\sum_{i=1}^{n} \frac{x_i x_{i'}}{\sum_{i''=1}^{n} x_{i''}^2} \right) y_{i'}.$$

so

$$a_i = \frac{x_i x_{i'}}{\sum_{i''=1}^{n} x_{i''}^2}$$