Q-1 (a)

$$E[(Y-X'b)^2] = \langle Y-X'b, Y-X'b\rangle$$

$$= \langle U+X'\beta_0-X'b, U+X'\beta_0-X'b\rangle$$

$$= \langle U+X'(\beta_0-b), U+X'(\beta_0-b)\rangle$$

$$= \langle U,U\rangle + \langle X'(\beta_0-b),U\rangle + \langle X'(\beta_0-b),U\rangle + \langle X'(\beta_0-b), X'(\beta_0-b)\rangle$$

$$= E[U^2] + 2\langle X'(\beta_0-b),U\rangle + \langle X'(\beta_0-b), X'(\beta_0-b)\rangle$$

$$= E[U^2] + 2E[(X'(\beta_0-b)U] + E[(X'(\beta_0-b))^2]$$

$$= E[U^2] + 2E[[(\beta_0-b)'XU)] + E[X'(\beta_0-b)X'(\beta_0-b)]$$

when $X'(\beta_0-b)U$ is scaler, so its transpose is equal to itself.

$(AB)' = B'A'$

Similarly, $\left(X'(\beta_0 - b)\right)' = (\beta_0 - b)'X$

$$E[(Y - X'b)] = E[U^2] + 2(\beta_0 - b)'E[XU] + (\beta_0 - b)'E[XX'](\beta_0 - b)$$

$\square$

(b)

We can write quadratic form,

$$(\beta_0 - b)'E[XX'](\beta_0 - b) = E[(\beta_0 - b)'XX'(\beta_0 - b)]$$
$$= E\left[\left[(\beta_0 - b)'X\right]^2\right]$$

It is given $E[(\beta_0 - b)'X)^2] \geq 0$ & is strictly positive if $\beta_0 \neq b$. So,

$$E[(Y-X'b)^2] = E[U^2] + 2(\beta_0-b)'E[XU] + E[((\beta_0-b)'X)^2]$$

$$= E[U^2] + 0 + \underbrace{E[((\beta_0-b)'X)^2]}_{\geq 0}$$

$$\geq E[U^2]$$

As we know, $E[((\beta_0-b)'X)^2] > 0$ if $\beta_0 \neq b$. So, strict inequality for $\beta_0 \neq b$.

As we see squared error of a linear predictor of Y is $(Y-X'b)^2$ and its expectation is $E[(Y-X'b)^2]$. Also, the expectation is minimized when $b = \beta_0$ i.e when we use population linear predictor.

(c)

$$V(Y) = E[(Y - E[Y])^2]$$

$$= E[(E^*[Y|X] + U - E[Y])^2]$$

$$= E[(E^*[Y|X] - E[Y] + U)^2]$$

$$= E\{(E^*[Y|X] - E[Y])^2 + 2(E^*[Y|X] - E[Y])(U) + U^2\}$$

Since, $E^*[Y|X]$ is projection, So projection error is orthogonal to $E^*[Y|X]$. Also, $E[U]=0$, So cross term is 0.

Also, $E[Y] = E[E^*[Y|X]]$

$$Var(Y) = E[(E^*[Y|X] - E[E^*[Y|X]])^2] + E[U^2]$$

$$Var(Y) = Var(E^*[Y|X]) + E[U^2]$$

but $E[U^2] = Var(Y - E^*[Y|X])$. So,

$$Var(Y) = Var(E^*[Y|x]) + Var(Y - E^*[Y|X])$$

So, total variance of Y variate due to linear component/linear projection plus the part not explained by linear projection.

(d)

Since X_K is scalar RV,

$$E^*[Y|X_K] = \alpha_K + \beta_K X_K$$

Since X_K, X_l are uncorrelated & $E^*[Y|X_K]$ is linear function of X_K,

$$Cov(E^*[Y|X_K], X_l) = 0$$

but since two are uncorrelated, projection of this onto X_ℓ,

$$E\left[E^*[Y|X_k]|X_\ell\right] = E\left[E^*[Y|X_k]\right] = E[Y]$$

$$E[UX_\ell] = E[YX_\ell] - \sum_{k=1}^{K} E\left[E^*[Y|X_k]|X_\ell\right] + (K-1)E[Y]E[X_\ell]$$

Since $E^*[Y|X_k]$ is lin,

$$E\left[E^*[Y|X_k]X_\ell\right] = E\left[E^*[Y|X_k]\right]E[X_\ell]$$
$$= E[Y]E[X_\ell]$$

for $k=\ell$,

$$E\left[E^*[Y|X_k]X_\ell\right] = E[YX_\ell]$$

$\Rightarrow$

$$E[UX_\ell] = E[YX_\ell] - E[YX_\ell] - (K-1)E[Y]E[X_\ell]$$
$$+(K-1)E[Y]E[X_\ell]$$

$$E[U X_l] = 0 + 0 = 0 \qquad \forall l$$

(e)

Note that when X_K is one regressor

then

$$E^*[Y \mid X_K] = \alpha_K + \beta_K X_K$$

where $\alpha_K = E[Y] - \beta_K E[X_K]$

$$\beta_K = \frac{Cov(Y, X_K)}{V(X_K)}$$

So,

$$E^*[Y \mid X_1, \dots X_N] = \sum_{k=1}^{N} E^*[Y \mid X_K] - (K-1) E[Y]$$

$$= E[Y] + \sum_{k=1}^{N} \left(E^*[Y \mid X_K] - E[Y] \right)$$

$$= E[Y] + \sum_{k=1}^{N} \left(\alpha_K + \beta_K X_K - E[Y] \right)$$

$$= E[Y] + \sum_{k=1}^{N} \left(E[Y] - \beta_k E[X_k] + \beta_k X_k - E[Y] \right)$$

$$= E[Y] + \sum_{k=1}^{K} \left(\beta_k \left(X_k - E[X_k] \right) \right)$$

$$= E[Y] + \sum_{k=1}^{K} \frac{Cov(Y, X_k)}{Var(X_k)} \left(X_k - E[X_k] \right)$$

Now,

$$Var(\hat{E}[Y | X_1, \dots X_K]) = Var\left(\sum_{k=1}^{K} \frac{C(Y, X_k)}{Var(X_k)} X_k \right)$$

as variance of expectation is 0.

$$= \sum_{k=1}^{K} \frac{C(Y, X_k)^2}{Var(X_k)^2} Var(X_k) +$$

$$+ \sum_{k \neq i} \frac{Cov(X, X_k) Cov(Y, X_i)}{Var(X_k) Var(Y_i)} C(X_k, Y_i)$$

$c(x_k, x_i) = 0$, so,

$$\mathrm{Var}(\hat{E}[Y|X_1, \dots X_N]) = \sum_{k=1}^{N} \frac{\mathrm{Cov}(Y, X_k)^2}{\mathrm{Var}(X_k)}$$

Now, $1 - \dfrac{\mathrm{Var}(U)}{\mathrm{Var}(Y)} = \dfrac{\mathrm{Var}(Y) - \mathrm{Var}(U)}{\mathrm{Var}(Y)}$

we showed in part (c),

$$\mathrm{Var}(Y) - \mathrm{Var}(U) = \mathrm{Var}(\hat{E}[Y|X])$$

$$1 - \frac{\mathrm{Var}(U)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(\hat{E}[Y|X_1, \dots X_N])}{\mathrm{Var}(Y)}$$

$$= \sum_{k=1}^{N} \frac{\mathrm{Cov}(Y, X_k)^2}{\mathrm{Var}(Y)\,\mathrm{Var}(X_k)}$$

$$= \sum_{k=1}^{N} \rho_k^2 \qquad \square$$

Q2-2 (a)

$$Y_i \sim N(\mu, \sigma_i^2)$$

choose $c = \begin{pmatrix} 1/N \\ 1/N \\ \vdots \\ 1/N \end{pmatrix}$ ( ]

then $c'Y = \frac{1}{N}Y_1 + \frac{1}{N}Y_2 + \cdots \frac{1}{N}Y_N = \bar{Y}$

So, there exists such a c.

(b)

$$MSE(\hat{u}) = Bias^2 + Var(\hat{u})$$

$$E[\hat{u}] = c' E[Y] = c' \mu \iota_N$$

$$Var(\hat{u}) = c' Var(Y) c = c' diag(\sigma_1^2, \ldots \sigma_N^2) c$$

$$Bias(\hat{u}) = \mu c' \iota_N - \mu \iota_N$$

$$MSE(\hat{u}) = c' \, diag(\sigma_1^2, \cdots \sigma_N^2) c + (\mu c' \iota_N - \mu)$$

Taking matrix derivative,

$$\frac{\partial MSE(\hat{u})}{\partial c} = 2\left(diag(\sigma_1^2, \cdots \sigma_N^2)\right) c + 2\mu \iota_N (\mu c' \iota_N - \mu) = 0$$

$$diag(\sigma_1^2, \cdots \sigma_N^2) c + \mu^2 \iota_N c' \iota_N - \mu^2 \iota_N = 0$$

$$\left[diag(\sigma_1^2, \cdots \sigma_N^2) + \mu^2 \iota_N \iota_N'\right] c = \mu^2 \iota_N$$

$$\boxed{c = \left[diag(\sigma_1^2, \cdots \sigma_N^2) + \mu^2 \iota_N \iota_N'\right]^{-1} \mu^2 \iota_N}$$

(d)

$\sigma_i^2 = \sigma^2.$

Let us call, $D = \mathrm{diag}(\sigma^2, \cdots \sigma^2)$

$$\iota_N' D^{-1} \iota_N = \sum_{i=1}^{N} \frac{1}{\sigma^2}$$

$$c = (D + \mathcal{H}^2 \iota_N \iota_N')^{-1} \mathcal{H}^2 \iota_N$$

$$(D + \mathcal{H}^2 \iota_N \iota_N')^{-1} = D^{-1} - \frac{\mathcal{H}^2 D^{-1} \iota_N \iota_N' D^{-1}}{1 + \mathcal{H}^2 \iota_N' D^{-1} \iota_N}$$

$$\frac{\mu^2 D^{-1} C_N C_N' D^{-1}}{H \mu^2 \sum_{i=1}^{N} 1/\sigma^2} =$$