1. (a) Let $Y = X\beta + \epsilon$ where $\epsilon \sim MVN(\mathbf{0}, \sigma^2 I)$. This means that $Y|X \sim MVN(X\beta, \sigma^2 I)$. From here we see that

$$\begin{aligned}\mathcal{L}(\beta) &= \det(2\pi\sigma^2 I)^{-1/2}\exp\Big\{-\frac{1}{2}(Y - X\beta)^T(\sigma^2 I)^{-1}(Y - X\beta)\Big\}\\ &= \det(2\pi\sigma^2 I)^{-1/2}\exp\Big\{-\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\Big\}\end{aligned}$$

Moreover, we can write the log-likelihood as follows

$$\begin{aligned}\log\mathcal{L}(\beta) &= -\frac{1}{2}\log(\det(2\pi\sigma^2 I)) - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)\\ &= -\frac{1}{2}\log(\det(2\pi\sigma^2 I)) - \frac{1}{2\sigma^2}\left[Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta\right]\end{aligned}$$

Differentiating with respect to β we get

$$\begin{aligned}\frac{d}{d\beta}\log\mathcal{L}(\beta) &= -\frac{1}{2\sigma^2}\left[-2Y^TY + 2\beta^TX^TX\right]\\ &= \frac{1}{\sigma^2}\left[Y^TX - \beta^TX^TX\right]\end{aligned}$$

Having calculated this quantity, we can now calculate its inner product.

$$\begin{aligned}\left(\frac{d}{d\beta}\log\mathcal{L}(\beta)\right)^T\left(\frac{d}{d\beta}\log\mathcal{L}(\beta)\right) &= \left(\frac{1}{\sigma^2}\left[Y^TX - \beta^TX^TX\right]\right)^T\left(\frac{1}{\sigma^2}\left[Y^TX - \beta^TX^TX\right]\right)\\ &= \frac{1}{\sigma^4}\left(X^TY - X^TX\beta\right)\left(Y^TX - \beta^TX^TX\right)\\ &= \frac{1}{\sigma^4}\big(X^TYY^TX - X^TY\beta^TX^TX\\ &- X^TX\beta Y^TX + X^TX\beta\beta^TX^TX\big)\end{aligned}$$

We are now ready to compute the Fisher information matrix.

$$\begin{aligned}\mathcal{I}(\beta) &= \mathbb{E}\Big[\frac{1}{\sigma^4}\big(X^TYY^TX - X^TY\beta^TX^TX - X^TX\beta Y^TX + X^TX\beta\beta^TX^TX\big)\Big]\\ &= \frac{1}{\sigma^4}\Big\{\mathbb{E}[X^TYY^TX] - \mathbb{E}[X^TY\beta^TX^TX] - \mathbb{E}[X^TX\beta Y^TX] + \mathbb{E}[X^TX\beta\beta^TX^TX]\Big\}\\ &= \frac{1}{\sigma^4}\Big\{\mathbb{E}[(X^TY)(X^TY)^T] - X^T\mathbb{E}(Y)\beta^TX^TX - X^TX\beta\mathbb{E}[Y^T]X + X^TX\beta\beta^TX^TX\Big\}\\ &= \frac{1}{\sigma^4}\Big\{\mathbb{E}[(X^TY)(X^TY)^T] - X^TX\beta\beta^TX^TX - X^TX\beta\beta^TX^TX + X^TX\beta\beta^TX^TX\Big\}\\ &= \frac{1}{\sigma^4}\Big\{\mathbb{V}[X^TY] + \mathbb{E}[X^TY]\mathbb{E}[X^TY]^T - X^TX\beta\beta^TX^TX\Big\}\\ &= \frac{1}{\sigma^4}\Big\{X^T\mathbb{V}[Y]X + (X^TX\beta)(X^TX\beta)^T - \beta^TX^TXX^TX\beta\Big\}\\ &= \frac{1}{\sigma^4}\Big\{\sigma^2X^TX + X^TX\beta\beta^TX^X - \beta^TX^TXX^TX\beta\Big\}\\ &= \frac{1}{\sigma^2}X^TX\end{aligned}$$

(b) Under the Fisher regularities, we can calculate the Fisher information matrix in the following way.

$$\begin{aligned}
\mathcal{I}(\beta) &= -\mathbb{E}\left[\frac{d^2\log\mathcal{L}(\beta)}{d\beta d\beta^T}\right]\\
&= -\mathbb{E}\left[\frac{d}{d\beta^T}\left(\frac{1}{\sigma^2}\left[Y^TX - \beta^TX^TX\right]\right)\right]\\
&= -\mathbb{E}\left[-\frac{1}{\sigma^2}X^TX\right]\\
&= \frac{1}{\sigma^2}X^TX
\end{aligned}$$

(c) From (a) and (b), we can establish the Cramer-Roa Lower Bound for the variance of an unbiased estimator of β as

$$[\mathcal{I}(\beta)]^{-1} = \left[\frac{1}{\sigma^2}X^TX\right]^{-1} = \sigma^2(X^TX)^{-1}$$

Now, from the standard linear model, we arrive at the estimator $\hat{\beta} = (X^TX)^{-1}X^TY$. Here, we see that

$$\mathbb{E}[\hat{\beta}|X] = (X^TX)^{-1}X^T\mathbb{E}[Y|X] = (X^TX)^{-1}X^TX\beta = \beta$$

which shows that this estimator $\hat{\beta}$ is unbiased for β. Moreover, we can calculate the variance in a similar way

$$\begin{aligned}
\mathbb{V}(\hat{\beta}|X) &= \mathbb{V}\big[(X^TX)^{-1}X^TY|X\big]\\
&= (X^TX)^{-1}X^T\mathbb{V}\big[Y|X\big]X(X^TX)^{-1}\\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}\\
&= \sigma^2(X^TX)^{-1}
\end{aligned}$$

Hence we see that $\hat{\beta}$ is an unbiased estimator that attains the Cramer-Roa lower bound and is therefore a UMVUE.

2. (a) Suppose that $Y = X\beta + \epsilon$ where the vector of errors is given by $\epsilon \sim MVN(0,\Sigma)$ where $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. As in problem one, we see that $Y|X \sim MVN(X\beta, \Sigma)$. From here we can write the likelihood as follows.

$$\mathcal{L}(\beta) = \det(2\pi\Sigma)^{-1/2}\exp\Big\{-\frac{1}{2}(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)\Big\}$$

where $\Sigma^{-1} = \text{diag}(1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_n^2)$.

(b) Using the likelihood above, we can compute the log-likelihood as follows.

$$\begin{aligned}\log \mathcal{L}(\beta) &= -1/2 \log[\det(2\pi\Sigma)] - 1/2(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)\\ &= -1/2 \log[\det(2\pi\Sigma)] - 1/2\Big\{Y^T\Sigma^{-1}Y - Y^T\Sigma^{-1}X\beta - \beta^T X^T\Sigma^{-1}Y + \beta^T X^T\Sigma^{-1}X\beta\Big\}\end{aligned}$$

Now differentiating with respect to β^T we get

$$\frac{d}{d\beta^T}\log \mathcal{L}(\beta) = -\frac{1}{2}\Big\{ - 2X^T\Sigma^{-1}Y + 2X^T\Sigma^{-1}X\beta\Big\}$$

Setting equal to zero and solving for $\hat{\beta}$ we get the following.

$$\begin{aligned}-\frac{1}{2}\Big\{ - 2X^T\Sigma^{-1}Y + 2X^T\Sigma^{-1}X\hat{\beta}\Big\} &= 0\\ X^T\Sigma^{-1}Y - X^T\Sigma^{-1}X\hat{\beta} &= 0\\ X^T\Sigma^{-1}X\hat{\beta} &= X^T\Sigma^{-1}Y\\ \hat{\beta} &= (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y\end{aligned}$$

Here we assumed that $X^T\Sigma^{-1}X$ is invertible (which occurs when X is full rank) and repeatedly used the fact that Σ was symmetric. This shows that that $\hat{\beta}$ is the same estimate found from using the weighted squared error loss function.