3. Multivariate linear regression is widely used to model the relationships between multiple related responses and a set of predictors. Suppose we have N observations of m-dimensional responses $Y_i = (y_{i,1}, \dots, y_{i,m})^\top$ and p-dimensional predictors $X_i = (x_{i,1}, \dots, x_{i,p})^\top$, for $i = 1, \dots, N$. Let $\mathbf{Y} = (Y_1, \dots, Y_N)^\top$ be the $N \times m$ response matrix and $\mathbf{X} = (X_1, \dots, X_N)^\top$ be the $N \times p$ covariates matrix. The multivariate linear regression model assumes

$$\mathbf{Y} = \mathbf{X}B + E,$$

where B is a $p \times m$ matrix of unknown regression parameters and $E = (\epsilon_1, \dots, \epsilon_N)^\top$ is an $N \times m$ matrix of regression errors, with ϵ_i's independently sampled from an m-dimensional Gaussian distribution $N(\mathbf{0}, \Sigma)$.

(a) For the multivariate linear regression model, derive the Maximum Likelihood Estimator (MLE) of B and Σ.

(b) Similarly to the unidimensional linear regression model, the Ordinary Least Squares (OLS) estimator of B is defined as the minimizer of the Residual Sum of Squares (RSS) defined as

$$RSS(B) = \sum_{i=1}^{N} (Y_i^\top - X_i^\top B)(Y_i^\top - X_i^\top B)^\top.$$

Derive the OLS estimator of B. Is it the same as the MLE?

(a) $L(B,\Sigma) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y_i - B^T X_i)^T \Sigma^{-1} (Y_i - B^T X_i)\right)$ — Likelihood function

$l(B,\Sigma) = \ln(B,\Sigma) = -\frac{Nm}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(Y_i - B^T X_i)^T \Sigma^{-1} (Y_i - B^T X_i)$

here $\sum_{i=1}^{N}(Y_i - B^T X_i)^T \Sigma^{-1}(Y_i - B^T X_i) = tr(\Sigma^{-1}(Y-XB)^T(Y-XB)) \Rightarrow l(B,\Sigma) = C - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}(Y-XB)^T(Y-XB))$

according to $\frac{d}{dB} tr(\Sigma^{-1}(Y-XB)^T(Y-XB)) = -2X^T(Y-XB)\Sigma^{-1}$, let $-2X^T(Y-XB)\Sigma^{-1} = 0 \Rightarrow X^T Y = X^T X B$

$\Rightarrow \hat{B}_{MLE} = (X^T X)^{-1} X^T Y$

define $S := (Y-XB)^T(Y-XB) \Rightarrow l(\Sigma) = -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}S)$

Here $\frac{dl}{d\Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}S = 0 \Rightarrow \hat{\Sigma}_{MLE} = \frac{1}{N}S = \frac{1}{N}(Y - X\hat{B}_{MLE})^T(Y - X\hat{B}_{MLE})$

(b) $RSS(B) = tr((Y-XB)^T(Y-XB)) = tr(Y^TY - Y^TXB - B^TX^TY + B^TX^TXB)$

$\frac{dRSS}{dB} = \frac{d}{dB} tr(-2Y^TXB + B^TX^TXB) = -2X^TY + 2X^TXB$, let $\frac{dRSS}{dB} = 0 \Rightarrow X^TXB = X^TY \Rightarrow \hat{B}_{OLS} = (X^TX)^{-1}X^TY$

$\hat{B}_{OLS} = \hat{B}_{MLE}$