Exercise 1, Expectation values for OLS expressions

$$E(Y_i) = E(f(X_i) + \varepsilon_i)$$

$$= E(f(X_i)) + E(\varepsilon_i) \rightarrow \varepsilon \sim N(0, \sigma^2)$$

non-stochastic

$$= f(X_i) = \tilde{Y}_i = \sum_j x_{ij}\beta_j = X_{i,*}\beta. \ \square$$

$$Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2$$

$$= E[(f(X_i) + \varepsilon_i)^2] - f(X_i)^2$$

$$= E[f(X_i)^2 + E(\varepsilon_i^2) + E(2f(X_i)\varepsilon_i)] - f(X_i)^2$$

$$= f(X_i)^2 + E(\varepsilon_i^2) - f(X_i)^2$$

$$= E(\varepsilon_i^2) = V(\varepsilon_i) + [E(\varepsilon_i)]^2$$

$$= \sigma^2. \ \square$$

$$E(\hat{\beta}) = E((X^TX)^{-1}X^TY) = \overbrace{(X^TX)^{-1}X^T}^{\text{non-stochastic}} E(Y)$$

$$= (X^TX)^{-1}X^TX\beta = \beta. \ \square$$

If X^TX is singular we can use pseudo-inverse via the SVD to show that $E(\hat{\beta}) = \beta$

This yields

$$(X^TX)^{-1}X^TX\beta = (V\Sigma^TU^TU\Sigma V^T)^{-1}(V\Sigma^TU^TU\Sigma V^T)\beta$$

$$= (V\Sigma^T\Sigma V^T)^{-1}(V\Sigma^T\Sigma V^T)\beta$$

$$= V(\tilde{\Sigma}^2)^{-1}V^TV\tilde{\Sigma}^2V^T\beta = \beta$$

$$Var(\hat{\beta}) = Var((X^TX)^{-1}X^Ty)$$

Set $H = (X^TX)^{-1}X^T$

$$Var(Hy) = H\,Var(y)\,H^T$$

where

$$Var(y) = Var(f(\vec{x}) + \varepsilon)$$
$$= Var(f(\vec{x})) + Var(\varepsilon) = \sigma^2$$

So $Var(Qy) = \sigma^2 HH^T$

$$HH^T = \underbrace{(X^TX)^{-1}X^TX}_{I}((X^TX)^{-1})^T$$

$$= ((X^TX)^{-1})^T = (X^TX)^{-1}$$

So $Var(\hat{\beta}) = \sigma^2(X^TX)^{-1}$. $\square$

Exercise 2, Expectation values for Ridge regression

$$E[\hat{\beta}_{Ridge}] = E((X^TX+\lambda I)^{-1}X^Ty))$$

$$= (X^TX+\lambda I)^{-1}X^TX\beta. \quad \square$$

$$Var[\hat{\beta}_{Ridge}] = Var[(X^TX+\lambda I)^{-1}X^Ty]$$

set $H = (X^TX+\lambda I)^{-1}X^T$

This yields

$$Var[Hy] = H\,Var(y)H^T$$

where $Var(y) = \sigma^2$

So $Var[Hy] = \sigma HH^T$

$$HH^T = (X^TX+\lambda I)^{-1}X^T\{(X^TX+\lambda I)^{-1}X^T\}^T$$

$$= (X^TX+\lambda I)^{-1}X^TX\{(X^TX+\lambda I)^{-1}\}^T$$

So $Var[\hat{\beta}_{Ridge}] = \sigma^2(X^TX+\lambda I)^{-1}X^TX\{(X^TX+\lambda I)^{-1}\}^T \quad \square$