Exercise 3.16 (find estimators of table 3.4)

Orthonormal $\underbrace{X^T X}_{I}$

LEAST SQUARES

$$\hat\beta^{OLS} = \underbrace{(X^T X)}_{I}^{-1} X^T y \rightarrow \hat\beta^{OLS} = X^T y \qquad \widehat{Var}(\hat\beta) = (X^T X)^{-1}\hat\sigma^2$$

BEST SUBSET (best M predictors)

Since orthonormality, $\hat\beta_j^{OLS} = x_j^T y$, $\forall j$ $s.e.(\hat\beta_j^{OLS}) = \hat\sigma$

$\Rightarrow$ the most relevant predictors are those with the highest $|\hat\beta_j^{OLS}|$

$$t = \frac{\hat\beta_j^{OLS} - 0}{s.e.(\hat\beta_j^{OLS})} \leftarrow \widehat{sd(\hat\beta^{OLS})}$$

$\Rightarrow$ best subset means selecting the predictors s.t. $|\hat\beta_j^{OLS}| \geq |\hat\beta_{(M)}^{OLS}|$ where (M) means the M-th largest $|\hat\beta_j^{OLS}|$

$$\Rightarrow \hat\beta_j^{best.subset} = \hat\beta_j^{OLS} \, \mathbb{1}\left(|\hat\beta_j^{OLS}| \geq |\hat\beta_{(M)}^{OLS}|\right)$$

RIDGE REGRESSION

$$\hat\beta^{ridge} = (X^T X + \lambda I)^{-1} X^T y$$

$$\overset{\text{orthonormal case}}{=} (I + \lambda I)^{-1} \underbrace{X^T y}_{\hat\beta^{OLS}}$$

Due to orthonormality, $\hat\beta_j^{ridge} = \dfrac{\hat\beta_j^{OLS}}{(1+\lambda)}$