Ex 3.16

Ridge: $\hat{\beta}^{ridge} = (X^TX + \lambda\mathbb{I})^{-1}X^Ty$

$= (\mathbb{I} + \lambda\mathbb{I})^{-1}X^Ty$ as $X^TX = \mathbb{I}$ when X orthonormal

notice, $\hat{\beta}^{ls} = (X^TX)^{-1}X^Ty$ (*)

$= X^Ty$

$\Rightarrow \hat{\beta}^{ridge} = \frac{\hat{\beta}^{ls}}{(\mathbb{I} + \lambda\mathbb{I})}$

e.g. $\hat{\beta}_j^{ridge} = \frac{\hat{\beta}_j^{ls}}{(1+\lambda)}$

Lasso: $\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}}\{\frac{1}{2}(y-X\beta)^T(y-X\beta) + \lambda\|\beta\|_1\}$

$\frac{\partial}{\partial\beta}(\frac{1}{2}(y-X\beta)^T(y-X\beta) + \lambda\|\beta\|_1) = 0$

$d\|X\|_1 = \text{Sign}(X)$

$\Rightarrow -X^Ty + X^TX\beta + \lambda\,\text{sign}(\beta) = 0$

$\hat{\beta}^{lasso} = X^Ty - \lambda\,\text{Sign}(\beta)$ $\quad X^TX = \mathbb{I}$

$= \hat{\beta}^{ls} - \lambda\,\text{Sign}(\hat{\beta}^{ls})$ using (*)

little bit more to do here...