Part a:

show that:

$$\hat{\beta} = X^T y$$

Is a closed form solution for the Ordinary Least Squares regression problem.

Reference: The Elements of Statistical Learning

Least squares pick β coefficients by minimizing the residual sum of squares (RSS)

$$\begin{aligned} RSS(\beta) &= \sum_{i=1}^{N}(y_i - f(x_i))^2 \\ &= \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{P} x_{ij}\beta_j\right)^2 \end{aligned}$$

Denote by X the $N \times (p+1)$ matrix with each row an input vector (with a 1 in the first position), and similarly let y be the N -vector of outputs in the training set. Then we can write the residual sum-of-squares as:

$$RSS = (y - X\beta)^T(y - X\beta)$$

This is a quadratic function in the $p+1$ parameters. Differentiating with respect to β we obtain:

$$\frac{\partial RSS}{\partial \beta} = -2X^T(y - X\beta)$$

$$\frac{\partial^2 RSS}{\partial \beta \partial \beta^T} = 2X^T X$$

Assuming (for the moment) that X has full column rank, and hence $X^T X$ is positive definite, we set the first derivative to zero:

$$X^T(y - X\beta) = 0$$

$$\text{Solve for } \beta\text{:}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Assuming X is orthonormal $X^TX = (X^TX)^{-1} = I^P$ where I^P is a $p \times p$ identity matrix. The fitted values at the training inputs are:

$$\begin{aligned}\hat{y} &= X(X^TX)^{-1}X^Ty = X\hat{\beta} \\ &= XIX^Ty = X\hat{\beta}\end{aligned}$$

Reducing the above and solving for $\hat{\beta}$ gives:

$$\hat{\beta} = X^Ty$$

Part b

Starting with the RSS in matrix form we have:

$$RSS(\lambda) = (y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$$

Differentiating with respect to then solving for β,

$$\begin{aligned}\frac{\partial RSS}{\partial\beta} &= -2X^T(y - X\beta) + 2\lambda\beta \\ \beta^{ridge} &= (X^TX + \lambda I)^{-1}X^Ty\end{aligned}$$

Since X is orthonormal and we already know from part a above that $\beta^{ols} = X^Ty$

$$\begin{aligned}\beta^{ridge} =& (I + \lambda I)^{-1}X^Ty \\ =& (1 + \lambda)^{-1}X^Ty \\ =& (1 + \lambda)^{-1}\beta^{ols}\end{aligned}$$

Part c

$$\begin{aligned}RSS(\lambda) =& (y - X\beta)^T(y - X\beta)\lambda|\beta| \\ \frac{\partial RSS}{\partial\beta} =& -2X^T(y - X\beta)\lambda\frac{\partial}{\partial\beta}|\beta| \\ 0 =& -2X^Ty2X^TX\beta\lambda\frac{\partial}{\partial\beta}|\beta| \\ 0 =& -2\beta^{ols}2I\beta\lambda\frac{\partial}{\partial\beta}|\beta| \\ 0 =& -2\beta_j^{Ols} + 2\beta_j + \lambda\frac{\partial}{\partial\beta}|\beta|\end{aligned}$$

The three cases on the sign for the partial derivative on the absolute value of β

$$0 = -\,2\beta_j^{Ols} + 2\beta_j + \lambda$$
$$0 = -\,2\beta_j^{Ols} + 2\beta_j - \lambda$$
$$0 = -\,2\beta_j^{Ols} + 2\beta_j$$

Which is the closed form solution:

$$\hat{\beta}_j^{lasso} = \begin{cases} \beta_j^{ols} - \frac{\lambda}{2} \text{ if } \beta_j^{ols} > frac\lambda 2 \\ 0 \text{ if } -\frac{\lambda}{2} \leq \beta_j^{ols} \leq \frac{\lambda}{2} \\ \beta_j^{ols} + \frac{\lambda}{2} \text{ if } \beta_j^{ols} < -\frac{\lambda}{2} \end{cases}$$

Didn't have time for part d, going camping for the holiday weekend :(