Exercise D

Suppose two exact copies of the same predictor are included, then:

$$\hat{\beta}_{ridge} = (X^TX+\lambda I)^{-1}X^Ty = \begin{bmatrix} \sum_{i=1}^{m} x_i^2 + \lambda & \sum_{i=1}^{m} x_i^2 \\ \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^2 + \lambda \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i y_i \end{bmatrix}$$

m (data are standardized) $m\,cov(x,y)$ (data are centered)

$$= \begin{bmatrix} m+\lambda & m \\ m & m+\lambda \end{bmatrix}^{-1} \begin{bmatrix} m\,cov(x,y) \\ m\,cov(x,y) \end{bmatrix}$$

m simplifies

$$= \begin{bmatrix} 1+\tilde{\lambda} & 1 \\ 1 & 1+ \end{bmatrix}^{-1} \begin{bmatrix} cov(x,y) \\ cov(x,y) \end{bmatrix}; \qquad \tilde{\lambda} = \frac{\lambda}{m}$$

$$= \begin{bmatrix} 1+\tilde{\lambda} & -1 \\ -1 & 1+\tilde{\lambda} \end{bmatrix} \frac{1}{(1+\tilde{\lambda})^2-1} \begin{bmatrix} cov(x,y) \\ cov(x,y) \end{bmatrix}$$

$$\Rightarrow \hat{\beta}_{1,ridge} = \hat{\beta}_{2,ridge} = \frac{(1+\tilde{\lambda})cov(x,y) - cov(x,y)}{\cancel{1} + 2\tilde{\lambda} + \tilde{\lambda}^2 \cancel{-1}} =$$

$$= \frac{cov(x,y)}{\tilde{\lambda}+2}.$$