Problem 4) Part 1:

$$\min_{\beta} : \|X\beta - y\|_2^2 = \sum (x_i \hat{\beta} - y_i)^2$$

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

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

Both are scalars and the transpose of a scalar is a scalar so $y^T X\beta = (y^T X\beta)^T = X^T \beta^T y$

$$= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta}$$

$$\frac{d}{d\hat{\beta}} LS = \frac{d}{d\hat{\beta}}\left(y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} \right)$$

set derivative to zero to optimize $\longrightarrow$

$$0 = -2X^T y + 2X^T X \hat{\beta}$$

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

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

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

Scanned by CamScanner

part 2 Ridge

$$\lambda \|\beta\|_2^2 = \lambda \beta^T \beta$$

$$\min_{\beta} : \|X\beta - y\|_2^2 + \lambda \|\beta\|_2^2$$

↓ using derivation from previous part

$$= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} + \lambda \beta^T \hat{\beta}$$

$$\frac{d}{d\beta} RR = -2X^T y + 2X^T X \hat{\beta} + 2\lambda \hat{\beta}$$

set to zero

$$0 = -2X^T y + 2X^T X \hat{\beta} + 2\lambda \hat{\beta}$$

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

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

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