a) OLS $\min_\beta \|y - X\beta\|_2^2$

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

$(y^T - \beta^T X^T)(y - X\beta) = y^T y - 2\beta^T X^T y + \beta^T \underbrace{\cancel{X^T X}}_{I \text{ orthogonality}} \beta$

$y^T y - 2\beta^T X^T y + \beta^T \beta$

take derivative WRT $\beta$

$0 = -2X^T y + 2\beta$ rearrange

$X^T y = \hat{\beta}^{OLS}$ ✓

b) Ridge $\min_\beta \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$

first part expands like before $y^T y - 2\beta^T X^T y + \beta^T \beta + \lambda \beta^2$

take derivative again $0 = -\cancel{2} X^T y + \cancel{2}\beta + \cancel{2}\lambda\beta$

$X^T y = (1+\lambda)\, \hat{\beta}^{Ridge}$

$\hat{\beta}^{Ridge} = \frac{1}{1+\lambda} \hat{\beta}^{OLS}$ ✓

c) Lasso $\min_\beta \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$

$= \underbrace{\cancel{y^T y}}_{\text{treated like constant}} - 2\beta^T X^T y + \beta^T \beta + \lambda |\beta|$ — not differentiable everywhere

$= -2\hat{\beta}_j^{OLS} \beta_j + \beta_j^2 + \lambda |\beta_j|$

$0 = -2\hat{\beta}_j^{OLS} + 2\beta_j + \lambda$

$\hat{\beta}_j^{Lasso} = \hat{\beta}_j^{OLS} - \lambda/2$

piece by piece

$\beta^{OLS} \leq \lambda/2$

$\beta^{OLS} > \lambda/2$ — when $\beta^{OLS} > 0$ then $-2\hat{\beta}^{OLS}\beta_j$ pos when $\beta_j < 0$

$\beta^{OLS} = 0 \rightarrow \beta_j = 0$ — when $\beta^{OLS} < 0$ then $-2\hat{\beta}^{OLS}\beta_j$ pos when $\beta_j > 0$

$\beta^{OLS} \geq -\lambda/2$

$\beta^{OLS} < -\lambda/2$

Scanned with CamScanner