1. $p=1$ 일 때

$$\hat{\beta}^{ridge} = \operatorname{argmin}\left(\|y - X\beta\|^2 + \lambda\|\beta\|^2 \right)$$

$$= \operatorname{argmin}\{ (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta \}$$

$$= \operatorname{argmin}\{ y^T y - \underbrace{y^T X\beta - \beta^T X^T y}_{-2\beta^T X^T y} + \beta^T X^T X \beta + \lambda \beta^T \beta \}$$

$$\frac{\partial \hat{\beta}^{ridge}}{\partial \beta} = -2(X^T y)^T + 2\beta^T X^T X + 2\lambda \beta^T = 0$$

$$2(X^T y)^T = 2\beta^T (X^T X + \lambda I)$$

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

$\beta = (X^T X + \lambda I)^{-1} (X^T y)$ $\rightarrow$ $p=1$ 이므로 $(X^T X + \lambda I)$는 scalar

$$= \frac{X^T y}{X^T X + \lambda}$$

2. $$\hat{\beta}^{ridge} = \operatorname{argmin}\{ \|y - x_1\beta_1 - x_2\beta_2\|^2 + \lambda\|\beta_1\|^2 + \lambda\|\beta_2\|^2 \}$$

i) $$\frac{\partial \hat{\beta}^{ridge}}{\partial \beta_1} = 2(y - x_1\beta_1 - x_2\beta_2)(-x_1) + 2\lambda\beta_1 = 0$$

$$x_1^T (y - x_1\beta_1 - x_2\beta_2) - \lambda\beta_1$$

$$x_1^T y - x_1^T x_1 \beta_1 - x_1^T x_2 \beta_2 - \lambda\beta_1 = 0$$

$$X^T y - X^T X \beta_1 - X^T X \beta_2 = \lambda \beta_1$$

ii) $$\frac{\partial \hat{\beta}^{ridge}}{\partial \beta_2} = x_2^T (y - x_1\beta_1 - x_2\beta_2) - \lambda\beta_2$$

$$= x_2^T y - x_2^T x_1 \beta_1 - x_2^T x_2 \beta_2 - \lambda\beta_2 = 0$$

$$X^T y - X^T X \beta_1 - X^T X \beta_2 = \lambda \beta_2$$

$$\beta_1 = \beta_2$$

$$X^T y - X^T X \beta_1 - X^T X \beta_1 = \lambda \beta_1$$

$$(2X^T X + \lambda) \beta_1 = X^T y$$

$$\beta_1 = \frac{X^T y}{2X^T X + \lambda}$$