$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} \quad X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \\ X_{n1} & \cdots & X_{np} \end{bmatrix}_{n\times p} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{p\times 1} \qquad Y = X\beta + \varepsilon$$

Ridge:

$$\begin{aligned} \hat\beta_\lambda = \frac{1}{n+\lambda} X^T Y &= \frac{1}{n+\lambda} X^T (X\beta + \varepsilon) \\ &= \frac{1}{n+\lambda} X^T X \beta + \frac{1}{n+\lambda} X^T \varepsilon \qquad (X^T X = n I_p) \\ &= \underbrace{\frac{n}{n+\lambda}\beta}_{\text{constant}} + \underbrace{\frac{1}{n+\lambda} X^T \varepsilon}_{\text{random}} \end{aligned}$$

Bias: (It is a $\mathbb{R}^p$ vector here)

$$\Rightarrow \mathbb{E}\hat\beta_\lambda = \frac{n}{n+\lambda}\beta + \frac{1}{n+\lambda}\underbrace{\mathbb{E}[X^T\varepsilon]}_{=0} = \frac{n}{n+\lambda}\cdot\beta \qquad ①$$

$$\text{Bias} = \mathbb{E}\hat\beta_\lambda - \beta = \frac{-\lambda}{n+\lambda}\beta \qquad (\text{if } \lambda = 0 \Rightarrow \text{OLS} .\ 0\text{-bias})$$

Cov: (It is a $p\times p$ matrix here)

$$\mathrm{Cov}(\hat\beta_\lambda) = \mathrm{Cov}\Big(\underbrace{\frac{n}{n+\lambda}\beta}_{\text{constant}} + \underbrace{\frac{1}{n+\lambda}}_{\text{R.V.}} X^T\varepsilon\Big) = \mathrm{Cov}\Big(\frac{1}{n+\lambda} X^T\varepsilon\Big)$$

For a random vector $U \in \mathbb{R}^p$, the cov-matrix is:

$$\operatorname{cov}(U) = \mathbb{E}\left[(U-\mathbb{E}U)(U-\mathbb{E}U)^T\right] \quad (p \times p \text{ matrix})$$

($U = \frac{1}{n+\lambda} X^T \varepsilon$. here, $\mathbb{E}\left[\frac{1}{n+\lambda} X^T\varepsilon\right] = 0$

$$\operatorname{Cov}\left(\frac{1}{n+\lambda} X^T\varepsilon\right) = \mathbb{E}\left[\left(\frac{1}{n+\lambda} X^T\varepsilon\right)\left(\frac{1}{n+\lambda} X^T\varepsilon\right)^T\right]$$

$$= \frac{1}{(n+\lambda)^2} \mathbb{E}\left[(X^T\varepsilon)(X^T\varepsilon)^T\right] = \frac{1}{(n+\lambda)^2} \mathbb{E}\left[X^T \varepsilon \varepsilon^T X\right]$$

$$= \frac{1}{(n+\lambda)^2} X^T \cdot \mathbb{E}[\varepsilon\varepsilon^T]\, X \overset{\mathbb{E}\varepsilon=0}{=\!=} \frac{1}{(n+\lambda)^2} X^T\, \mathbb{E}\left[(\varepsilon-\mathbb{E}\varepsilon)(\varepsilon-\mathbb{E}\varepsilon)^T\right] X$$

$$= \frac{1}{(n+\lambda)^2} X^T \underbrace{\operatorname{Cov}(\varepsilon)}_{\substack{\parallel \\ \sigma^2 I_n \text{ by def}}} X = \frac{\sigma^2}{(n+\lambda)^2} X^T X = \frac{n\sigma^2}{(n+\lambda)^2} \cdot I_p \qquad 2$$

Thus, $\operatorname{cov}(\hat\beta_\lambda) = \operatorname{cov}\left(\frac{1}{n+\lambda} X^T\varepsilon\right) = \frac{n \cdot \sigma^2}{(n+\lambda)^2} \cdot I_p$

($\lambda = 0 \Rightarrow$ OLS cov-matrix)

Error: (It is a scalar here)

$$\mathbb{E}\|\hat\beta_\lambda - \beta\|_2^2 = \mathbb{E}\,(\hat\beta_\lambda - \beta)^T(\hat\beta_\lambda - \beta)$$

$$= \mathbb{E} \sum_{i=1}^{p} (\hat\beta_{\lambda,i} - \beta_i)^2$$

$$= \sum_{i=1}^{p} \mathbb{E}\left[(\hat\beta_{\lambda,i} - \beta_i)^2\right]$$

$$\hat\beta_\lambda = \begin{bmatrix} \hat\beta_{\lambda,1} \\ \vdots \\ \hat\beta_{\lambda,p} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Using bias-var decomposition:

$$\hat{\beta}_{\lambda,i}-\beta_i = (\hat{\beta}_{\lambda,i}-\mathbb{E}\hat{\beta}_{\lambda,i})+(\mathbb{E}\hat{\beta}_{\lambda,i}-\beta_i)$$

$$\mathbb{E}[(\hat{\beta}_{\lambda,i}-\beta_i)^2] = \underbrace{\mathbb{E}[(\hat{\beta}_{\lambda,i}-\mathbb{E}\hat{\beta}_{\lambda,i})^2]}_{\text{Var of } \hat{\beta}_{\lambda,i}} + \underbrace{(\mathbb{E}\hat{\beta}_{\lambda,i}-\beta_i)^2}_{\text{Bias}^2 \text{ of } \hat{\beta}_{\lambda,i}}$$

By the definition of cov-matrix

Var of $\hat{\beta}_{\lambda,i}$ is the (i,i)-entry of cov-matrix in ②

$$= \frac{n\sigma^2}{(n+\lambda)^2}$$

Bias^2 of $\hat{\beta}_{\lambda,i}$ is the square of ith element of bias vector in ①

$$= \frac{\lambda^2}{(n+\lambda)^2}\beta_i^2$$

$$\Rightarrow \mathbb{E}(\|\hat{\beta}_\lambda-\beta\|_2^2) = \sum_{i=1}^{p}\mathbb{E}(\hat{\beta}_{\lambda,i}-\beta_i)^2$$

$$= \sum_{i=1}^{p}(\text{Var of } \hat{\beta}_{\lambda,i} + \text{Bias}^2 \text{ of } \hat{\beta}_{\lambda,i})$$

$$= \sum_{i=1}^{p}\left(\frac{n\sigma^2}{(n+\lambda)^2}+\frac{\lambda^2}{(n+\lambda)^2}\beta_i^2\right) = \frac{np\sigma^2}{(n+\lambda)^2}+\frac{\lambda^2}{(n+\lambda)^2}\underbrace{\sum_{i=1}^{p}\beta_i^2}_{=\|\beta\|_2^2}$$

In matrix form, we have

$$\mathbb{E}\|\hat{\beta}_\lambda-\beta\|_2^2 = \mathbb{E}[(\hat{\beta}_\lambda-\beta)^T(\hat{\beta}_\lambda-\beta)]$$

$$\overset{(*)}{=} \mathbb{E}[tr\{(\hat{\beta}_\lambda-\beta)^T(\hat{\beta}_\lambda-\beta)\}]$$

$$= \mathbb{E}[tr\{(\hat{\beta}_\lambda-\beta)(\beta_\lambda-\beta)^T\}]$$

$$= \mathbb{E}[tr\{((\hat{\beta}_\lambda-\mathbb{E}\hat{\beta}_\lambda)+(\mathbb{E}\hat{\beta}_\lambda-\beta))((\hat{\beta}_\lambda-\mathbb{E}\hat{\beta}_\lambda)+(\mathbb{E}\hat{\beta}_\lambda-\beta))^T\}]$$

$$\overset{(\Delta)}{=} tr\{\mathbb{E}[(\hat{\beta}_\lambda-\mathbb{E}\hat{\beta}_\lambda)(\hat{\beta}_\lambda-\mathbb{E}\hat{\beta}_\lambda)^T] + (\mathbb{E}\hat{\beta}_\lambda-\beta)(\mathbb{E}\hat{\beta}_\lambda-\beta)^T\}$$

$$\overset{(**)}{=} tr(Cov(\hat{\beta}_\lambda)) + tr((\mathbb{E}\hat{\beta}_\lambda-\beta)^T(\mathbb{E}\hat{\beta}-\beta))$$

$$= \frac{np\sigma^2}{(n+\lambda)^2} + \frac{\lambda^2}{(n+\lambda)^2}\|\beta\|_2^2$$

In $(*)$ and $(**)$ we use

$$u^Tv = tr(u^Tv) = tr(vu^T) \quad (u,v\in\mathbb{R}^p)$$

In (Δ) we use $tr(A+B) = tr(A) + tr(B)$

$A,B\in\mathbb{R}^{p\times p}$ (matrix)