Derivation of the bias-variance decomposition

$$\mathbb{E}_{\mathcal{Z}} \|\hat{w} - w^*\|^2 = \mathbb{E}_{\mathcal{Z}} \|\hat{w} - \bar{w} + \bar{w} - w^*\|^2 \quad (2)$$

$$= \mathbb{E}_{\mathcal{Z}} \left(\|\hat{w} - \bar{w}\|^2 + 2(\hat{w} - \bar{w})^T (\bar{w} - w^*) + \|\bar{w} - w^*\|^2 \right)$$

(1)

$$= \underbrace{\boxed{\mathbb{E}_{\mathcal{Z}} \|\hat{w} - \bar{w}\|^2}}_{\text{Var}} + \cancel{2\mathbb{E}_{\mathcal{Z}} (\hat{w} - \bar{w})^T (\bar{w} - w^*)} \underbrace{\boxed{+ \|\bar{w} - w^*\|^2}}_{\text{Bias}^2}$$

$$\mathbb{E}_{\mathcal{Z}}\, \hat{w} - \bar{w} = 0$$

Derivation of the bias

$$\|\bar{w} - w^*\|^2 = \left\| \frac{\lambda I}{X^T X + \lambda I} w^* \right\|^2$$

$$\|\bar{w} - w^*\|^2$$

$$= (X^T X + \lambda I_p)^{-1} \left(\cancel{X^T X} w^* - (\cancel{X^T X} + \lambda I_p) w^* \right)$$

$$= \| -\lambda (X^T X + \lambda I_p)^{-1} w^* \|^2$$

$$\text{①}$$

$$= \lambda^2 w^{*T} (X^T X + \lambda I_p)^{-2} w^*$$

$$= \lambda^2 w^{*T} V (S^2 + \lambda I_p)^{-2} V^T w^*$$

$$= \lambda^2 w^{*T} \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} \begin{bmatrix} \frac{1}{(S_1^2+\lambda)^2} & & O \\ & \ddots & \\ O & & \frac{1}{(S_p^2+\lambda)^2} \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_p^T \end{bmatrix} w^*$$

$$= \lambda^2 \begin{bmatrix} w^{*T} v_1 & \cdots & w^{*T} v_p \end{bmatrix} \begin{bmatrix} \cdot & & \\ & \ddots & \\ & & \ddots \end{bmatrix} \begin{bmatrix} v_1^T w^* \\ \vdots \\ v_p^T w^* \end{bmatrix}$$

$$\bar{w} = (X^T X + \lambda I)^{-1} X^T X w^*$$

$$\sum_{j=1}^{p} \lambda^2 \frac{w^{*T} v_j \, v_j^T w^*}{(S_j^2 + \lambda)^2}$$

$$= \sum_{j=1}^{p} \left(\frac{\lambda v_j^T w^*}{S_j^2 + \lambda} \right)^2$$

Derivation of the variance

$$\hat{w} - \bar{w} = (X^T X + \lambda I_p)^{-1} X^T \xi$$

$$E_\xi \|\hat{w} - \bar{w}\|^2 = E_\xi \underbrace{\boxed{\xi^T X}}_{A} \underbrace{\boxed{(X^T X + \lambda I_p)^{-2} X^T \xi}}_{B} \qquad 1$$

$$= E_\xi \, Tr\left[(X^T X + \lambda I_p)^{-2} X^T \xi \xi^T X\right] \qquad 5$$

$$= Tr\left[(X^T X + \lambda I_p)^{-2} X^T \underbrace{E_\xi(\xi \xi^T)}_{\sigma^2 I} X\right] \qquad 11$$

$$= \sigma^2 Tr\left[(X^T X + \lambda I_p)^{-2} X^T X\right] \qquad 13$$

$$= \sigma^2 Tr\left[\underbrace{\boxed{V (S^2 + \lambda I_p)^{-2} \underbrace{\cancel{V^T V}}_{I} S^2}}_{A} \underbrace{\boxed{V^T}}_{B}\right] \qquad 5$$

$$= \sigma^2 Tr\left[(S^2 + \lambda I_p)^{-2} S^2\right]$$

$$= \sigma^2 \sum_{i=1}^{m} \frac{s_i^2}{(s_i^2 + \lambda)^2}$$

$$Tr(AB) = Tr(BA)$$

$AB \quad m \times m$

$A \quad m \times k$

$B \quad k \times m$

$A \quad 1 \times p$

$B \quad p \times 1$

$$\xi \sim N(0, \sigma^2 I)$$

$$X = USV^T$$

- $X^T X = V S^2 V^T$
- $X^T X + \lambda I_p = V (S^2 + \lambda I_p) V^T$

First step for the analysis of lasso

$$\frac{1}{2n}\|y - X\hat{w}\|^2 + \lambda_n \|\hat{w}\|_1 \le \frac{1}{2n}\|y - Xw^*\|^2 + \lambda_n \|w^*\|_1$$

$$y = Xw^* + \xi$$

$$\frac{1}{2n}\|X(w^* - \hat{w}) + \xi\|^2 + \lambda_n \|\hat{w}\|_1 \le \frac{1}{2n}\|\xi\|^2 + \lambda_n \|w^*\|_1 \quad ②$$

$$\frac{1}{2n}\|X(w^* - \hat{w})\|^2 + \frac{1}{n}\xi^T X(w^* - \hat{w}) + \cancel{\frac{1}{2n}\|\xi\|^2} \le \cancel{\frac{1}{2n}\|\xi\|^2} + \lambda_n \left(\|w^*\|_1 - \|\hat{w}\|_1\right)$$

$$\frac{1}{2n}\|X(w^* - \hat{w})\|^2 \le \frac{1}{n}\xi^T X(\hat{w} - w^*) + \lambda_n \left(\underbrace{\|w^*\|_1 - \|\hat{w}\|_1}_{\le \|w^* - \hat{w}\|_1}\right)$$

$$\le \|X^T\xi/n\|_\infty \|\hat{w} - w^*\|_1 + \lambda_n \|w^* - \hat{w}\|_1$$

Bound on the inf-norm of input-noise correlation

$$z_j \sim N(0, \underline{\sigma^2 \|x_j\|^2})$$

$$\underline{\max \sigma_j} = \sigma \max_j \|x_j\| = \underline{\sigma \sqrt{n} R} \qquad R := \frac{\max_j \|x_j\|}{\sqrt{n}}$$

$$\Pr\Big(\underbrace{\max_j |z_j|}_{\|X^T z\|_\infty} > 2\underline{\sigma\sqrt{n}R}\sqrt{\log p}\Big) \le \frac{2}{p}$$

$$\Pr\Big(\|X^T z\|_\infty / n \geq 2R\sqrt{\frac{\log p}{n}}\Big) \le \frac{2}{p}$$

Derivation of the "better bound"

$$\|w\|_1 = \sum_{j=1}^{p} |w_j| = x^T w \qquad x = \mathrm{sign}(w) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ +1 \\ -1 \\ \vdots \\ 0 \end{bmatrix} \; k$$

$$\le \|x\|_2 \, \|w\|_2$$

$$= \sqrt{K} \, \|w\|_2$$

$$\|\Delta\|_1 = \|\Delta'\|_1 + \|\Delta''\|_1 \qquad \Delta = \hat{w} - w^*$$

$$\|w^*\|_1 - \|\hat{w}\|_1$$

$$= \|w^*\|_1 - \|\Delta + w^*\|_1 \qquad \Delta = \Delta' + \Delta''$$

$$= \|w^*\|_1 - \|\boxed{w^* + \Delta'} + \boxed{\Delta''}\|_1 \qquad \|w^* + \Delta'\|_1 \ge \|w^*\|_1 - \|\Delta'\|_1$$

$$= \|w^*\|_1 - \left(\|w^* + \Delta'\|_1 + \|\Delta''\|_1\right)$$

$$\le \cancel{\|w^*\|_1} - \cancel{\|w^*\|_1} + \|\Delta'\|_1 - \|\Delta''\|_1 = \|\Delta'\|_1 - \|\Delta''\|_1$$

Bounding the non-sparse part

$$0 \le \|X(\hat{w} - w^*)\|^2 \le \underbrace{\|X^T \xi / n\|_\infty}_{\le \frac{\lambda_n}{2}} \|\hat{w} - w^*\|_1 + \lambda_n \underbrace{\left(\|w^*\|_1 - \|\hat{w}\|_1\right)}_{\le \|\Delta^{S}\|_1 - \|\Delta^{S^c}\|_1}$$

$$0 \le \frac{\lambda_n}{2} \underbrace{\|\hat{w} - w^*\|_1}_{\Delta^{S} + \Delta^{S^c}} + \lambda_n \left(\|\Delta^{S}\|_1 - \|\Delta^{S^c}\|_1\right)$$

$$= \frac{\lambda_n}{2}\left(\|\Delta^{S}\|_1 + \|\Delta^{S^c}\|_1\right) + \lambda_n\left(\|\Delta^{S}\|_1 - \|\Delta^{S^c}\|_1\right)$$

$$= \frac{3}{2}\lambda_n \|\Delta^{S}\|_1 - \frac{1}{2}\lambda_n \|\Delta^{S^c}\|_1$$

$$\|\Delta^{S^c}\|_1 \le 3\|\Delta^{S}\|_1$$

$$\|\Delta\|_1 = \|\Delta^{S}\|_1 + \|\Delta^{S^c}\|_1 \le 4\|\Delta^{S}\|_1 \le 4\sqrt{k}\,\|\Delta^{S}\|_2 \le 4\sqrt{k}\,\|\Delta\|_2$$

Derivation of the lower-bound

$$\frac{1}{\sqrt{n}} \| X(\hat{w} - w^*) \|_2 \geq \frac{1}{4} \| \hat{w} - w^* \|_2 - 9 \sqrt{\frac{\log p}{n}} \| \hat{w} - w^* \|_1$$

$$\leq 4\sqrt{k} \| \hat{w} - w^* \|_2$$

$$\geq \left(\frac{1}{4} - 36 \left(\sqrt{\frac{k \log p}{n}} \right) \right) \| \hat{w} - w^* \|_2$$

$$n \geq c_1 k \log p$$

$$\geq \left(\frac{1}{4} - \frac{36}{\sqrt{c_1}} \right) \| \hat{w} - w^* \|_2$$