Fundamentals of Machine Learning - Exercise 6

Dominique Cheray & Hanuel Krämer

1 Bias & Variance of ridge regression

$$\hat{\beta}_\tau = \underset{\beta}{\operatorname{argmin}} \|X\beta - Y\|_F^2 + \tau \|\beta\|_2^2$$

$$\hat{\beta}_\tau = \left(X^T X + \tau \mathbb{1}\right)^{-1} X^T Y$$

$$Y = X\beta^* + \varepsilon$$

$$\mathbb{E}\left[\hat{\beta}_\tau\right] = \mathbb{E}\left[\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T Y\right] = \mathbb{E}\left[\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T X \beta^*\right] + \mathbb{E}\left[\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon\right] =$$

$$= \mathbb{E}\left[S_\tau^{-1} \cdot S \cdot \beta^*\right] + \mathbb{E}\left[S_\tau^{-1} \cdot X^T \cdot \varepsilon\right] = S_\tau^{-1} S \beta^* + S_\tau^{-1} X^T \cdot \underbrace{\mathbb{E}[\varepsilon]}_{=0} = S_\tau^{-1} S \beta^*$$

$$\mathbb{E}\left[\left(\hat{\beta}_\tau - \mathbb{E}[\hat{\beta}_\tau]\right) \cdot \left(\hat{\beta}_\tau - \mathbb{E}[\hat{\beta}_\tau]\right)^T\right] = \mathbb{E}\left[\left(\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T Y - S_\tau^{-1} S \beta^*\right)\left(\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T Y - S_\tau^{-1} S \beta^*\right)^T\right] =$$

$$= \mathbb{E}\left[\left(\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T X \beta^* + \left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon - S_\tau^{-1} S \beta^*\right)\left(\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T X \beta^* + \left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon - S_\tau^{-1} S \beta^*\right)^T\right] =$$

$$= \mathbb{E}\left[\left(S_\tau^{-1} S \beta^* + \left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon - S_\tau^{-1} S \beta^*\right)\left(S_\tau^{-1} S \beta^* + \left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon - S_\tau^{-1} S \beta^*\right)^T\right) =$$

$$= \mathbb{E}\left[\left(\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T \varepsilon\right)\left(\underbrace{\left(X^T X + \tau \mathbb{1}\right)^{-1}}_{\left(\left(X^T X + \tau \mathbb{1}\right)^{-1}\right)^T} X^T \varepsilon\right)^T\right] = \mathbb{E}\left[\left(X^T X + \tau \mathbb{1}\right)^{-1}\left(X^T X + \tau \mathbb{1}\right)^{-1} X^T X \varepsilon \varepsilon^T\right] =$$

$$= \left(\left(X^T X + \tau \mathbb{1}\right)^T\right)^{-1} \qquad \left(X^T X + \tau \mathbb{1}\right)^T = \left(X^T X + \tau \mathbb{1}\right)$$

$$= \left(X^T X + \tau \mathbb{1}\right)^{-1}$$

$$= S_\tau^{-2} \cdot S \cdot \underbrace{\mathbb{E}\left[\varepsilon \varepsilon^T\right]}_{=\sigma^2} = S_\tau^{-2} S \sigma^2$$