Fundamentals of Machine Learning - Exercise 6

Eugen Dizer
Luis Pazos Clemens
Robert Freund

2. Bias and variance of ridge regression

From: $\hat{\beta}_\tau = \arg\min_\beta \|X\beta - y\|_2^2 + \tau\|\beta\|_2^2$

we derived in the lecture: $\hat{\beta}_\tau = (X^T X + \tau \mathbb{1})^{-1} X^T y$

insert: $y = X\beta^* + \varepsilon \Rightarrow \hat{\beta}_\tau = S_\tau^{-1} S \beta^* + S_\tau^{-1} X^T \varepsilon$

using: $S = X^T X$ and $S_\tau = X^T X + \tau \mathbb{1}$

We want to calculate: $\mathbb{E}[\hat{\beta}_\tau] = \mathbb{E}[S_\tau^{-1} S \beta^*] + \mathbb{E}[S_\tau^{-1} X^T \varepsilon]$

because β^* is a constant and noise ε is indep from X:

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

$$= \mathbb{E}[S_\tau^{-1} S]\beta^*$$ because $\varepsilon \propto N(0, \sigma^2)$

for one TS $\rightarrow$ $= S_\tau^{-1} S \beta^*$

Covariance: $\mathrm{Cov}[\hat{\beta}_\tau] = \mathbb{E}[(\hat{\beta}_\tau - \mathbb{E}[\hat{\beta}_\tau])^2]$

We again calculate the expectation value by using one TS:

$$= \mathbb{E}[(S_\tau^{-1} S \beta^* + S_\tau^{-1} X^T \varepsilon - S_\tau^{-1} S \beta^*)^2]$$

$$= \mathbb{E}[S_\tau^{-1} X^T \varepsilon \varepsilon^T X S_\tau^{-1}]$$ using $[S_\tau^{-1}]^T = S_\tau^{-1}$ because S_τ is symmetric

$$= \mathbb{E}[S_\tau^{-1} X^T]\, \mathbb{E}[\varepsilon \varepsilon^T]\, \mathbb{E}[X S_\tau^{-1}]$$

$\hookrightarrow$ using ε is indep. of X (like in lecture) and

$$\sigma^2 = \mathbb{E}[\varepsilon\varepsilon^T]$$

$$\mathrm{Cov}[\hat{\beta}_\tau] = \sigma^2 \mathbb{E}[S_\tau^{-1} X^T X S_\tau^{-1}] = \sigma^2 \mathbb{E}[S_\tau^{-1} S S_\tau^{-1}]$$

$$= \sigma^2 S_\tau^{-1} S S_\tau^{-1}$$