Machine Learning - Exercise 6

Johannes Kammerer, Zhao Sun, Tong Yu

February 3, 2021

2 Bias and variance of ridge regression (8 points)

We know:

$$
\begin{aligned}
y &= X\beta^* + \epsilon && (1)\\
\hat{\beta} &= (X^T X + \tau 1)^{-1} X^T y && (2)\\
&= S_\tau^{-1} X^T y && (3)\\
&= S_\tau^{-1} S\beta^* + S_\tau^{-1} X^T \epsilon && (4)
\end{aligned}
$$

Then, analogous to the case $\tau = 0$ in the lecture:

$$
\begin{aligned}
E(\hat{\beta}) &= E(S_\tau^{-1} S\beta^* + S_\tau^{-1} X^T \epsilon) && (5)\\
&= E(S_\tau^{-1} S\beta^*) + E(S_\tau^{-1} X^T) E(\epsilon) && (6)\\
&= S_\tau^{-1} S\beta^*, && (7)
\end{aligned}
$$

because $\epsilon \sim N(0, \sigma^2)$, hence $E(\epsilon) = 0$.

Similar for the covariance:

$$
\begin{aligned}
cov[\hat{\beta}_\tau] &= E[(\hat{\beta}_\tau - E(\hat{\beta}_\tau))^2] && (8)\\
&= E[(\hat{\beta}_\tau - S_\tau^{-1} S\beta^*)^2] && (9)\\
&= E[(S_\tau^{-1} S\beta^* + S_\tau^{-1} X^T \epsilon - S_\tau^{-1} S\beta^*)^2] && (10)\\
&= E[(S_\tau^{-1} X^T \epsilon)(S_\tau^{-1} X^T \epsilon)^T] && (11)\\
&= E(S_\tau^{-1} X^T \epsilon\epsilon^T X S_\tau^{-1}) && (12)\\
&= E(S_\tau^{-1} S S_\tau^{-1}) E(\epsilon\epsilon^T) && (13)\\
&= S_\tau^{-1} S S_\tau^{-1} \sigma^2 && (14)
\end{aligned}
$$

QED.