QBUS6810 Statistical Learning and Data Mining

Tutorial 2 (Written Exercise)

Decompose the Expected Prediction Error into three parts: Irreducible error, squared Bias, and Variance (slides 33-35 of Lecture 1).

Solution: We treat $\boldsymbol{x}_0$ and f as fixed, i.e. non-random. Note that Y_0 is not in the original sample, and so $\widehat{f}$ is independent of both Y_0 and ε_0. We have:

$$\begin{aligned}\text{EPE} &= E\left[\left(Y_0 - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(f(\boldsymbol{x}_0) + \varepsilon_0 - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(\varepsilon_0 + \left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right)^2\right] \\ &= E\left[\varepsilon_0^2\right] + 2E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] + E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right].\end{aligned}$$

Note that $2E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] = 0$, because ε_0 is independent from $\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)$, and $E\left[\varepsilon_0\right] = 0$, which implies $E\left[\varepsilon_0\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)\right] = E\left[\varepsilon_0\right] E\left[f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right] = 0$. Also note that, by definition, $\text{Var}\left(\varepsilon_0\right) = E\left[\varepsilon_0^2\right] - \left(E\left[\varepsilon_0\right]\right)^2$. Thus, $E\left[\varepsilon_0^2\right] = \text{Var}\left(\varepsilon_0\right) = \sigma^2$. Consequently,

$$\text{EPE} = \sigma^2 + E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = \text{Irreducible error} + \text{Reducible error}.$$

Now we focus on the Reducible error:

$$\begin{aligned}&E\left[\left(f(\boldsymbol{x}_0) - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] = E\left[\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right] + E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] \\ &= \left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2 + 2E\left[\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)\right] \\ &+ E\left[\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right].\end{aligned}$$

The middle term is zero again. To see this note that the only random component in this term is $\widehat{f}(\boldsymbol{x}_0)$, and $E\left[E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right] = 0$. Hence, the Reducible error is:

$$\begin{aligned}&\left(f(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2 + E\left[\left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - \widehat{f}(\boldsymbol{x}_0)\right)^2\right] \\ &= \left(E\left[\widehat{f}(\boldsymbol{x}_0)\right] - f(\boldsymbol{x}_0)\right)^2 + E\left[\left(\widehat{f}(\boldsymbol{x}_0) - E\left[\widehat{f}(\boldsymbol{x}_0)\right]\right)^2\right] \\ &= \text{Bias}^2\left(\widehat{f}(\boldsymbol{x}_0)\right) + \text{Var}\left(\widehat{f}(\boldsymbol{x}_0)\right)\end{aligned}$$

Putting it all together,

$$\text{EPE} = \sigma^2 + \text{Bias}^2\left(\widehat{f}(\boldsymbol{x}_0)\right) + \text{Var}\left(\widehat{f}(\boldsymbol{x}_0)\right).$$