SHEET - An Introduction to Statistical Learning
Chapter 2 - Statistical Learning

1 December 2023

1 Statistical Learning

We suppose that we observe a quantitative response Y and p different predictors, $X_1, X_2, \ldots, X_p$. We assume that there is some relationship between Y and $X = (X_1, X_2, \ldots, X_p)$, which can be written in the very general form

$$Y = f(X) + \varepsilon.$$

Here f is some fixed but unknown function of $X_1, \ldots, X_p$, and ε is a random error term, which is independent of X and has mean zero. In this formulation, f represents the systematic information that X provides about Y. Consider a given estimate $\hat{f}$ and a set of predictors X, which yields the prediction $\hat{Y} = \hat{f}(X)$. Assume for a moment that both $\hat{f}$ and X are fixed. Then, we show that

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(X) + \varepsilon - \hat{f}(X)]^2 \\ &= [f(X) - \hat{f}(X)]^2 + \mathrm{Var}(\varepsilon) \qquad (2.3) \\ &= \mathrm{Var}(\hat{f}) + \mathrm{Bias}(\hat{f})^2 + \mathrm{Var}(\varepsilon) \qquad (2.7) \end{aligned}$$

Demonstration: $E(y - \hat{y})^2$

We have

$$\begin{aligned} E[\varepsilon] &= 0 \\ E[y] &= E[f + \varepsilon] = E[f] = f \\ Var[X] &= E[X^2] - E[X]^2 \Leftrightarrow E[X^2] = \mathrm{Var}[X] + E[X]^2 \end{aligned}$$

Then

$$\begin{aligned} \mathbf{E}[(\mathbf{y} - \hat{\mathbf{y}})^\mathbf{2}] &= E[(y - \hat{f})^2] = E[y^2] + E[\hat{f}^2] - E[2y\hat{f}] \\ &= \mathrm{Var}[y] + E[y]^2 + \mathrm{Var}[\hat{f}] + E[\hat{f}]^2 - 2fE[\hat{f}] \\ &= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + (f^2 - 2fE[\hat{f}] + E[\hat{f}]^2) \end{aligned}$$

We have

$$\begin{aligned} (E[f^2 - \hat{f}])^2 = (f^2 - E[\hat{f}])^2 &= f^2 - 2fE[\hat{f}] + E[\hat{f}]^2 \\ Var(y) = E[(y - E[y])^2] = E[(f + \varepsilon - f)^2] &= E[\varepsilon^2] = Var(\varepsilon) \end{aligned}$$

Then

$$\begin{aligned} \mathbf{E}[(\mathbf{y} - \hat{\mathbf{y}})^\mathbf{2}] &= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + (f - E[\hat{f}])^2 \\ &= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + E[(f - \hat{f})]^2 \\ \mathbf{E}[(\mathbf{y} - \hat{\mathbf{y}})^\mathbf{2}] &= \mathbf{Var}[\varepsilon] + \mathbf{Var}[\hat{\mathbf{f}}] + \mathbf{Bias}[\hat{\mathbf{f}}]^\mathbf{2} \end{aligned}$$