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

PLAYE Nicolas

1 December 2023

1 Statistical Learning

1.1 Courses' Demonstrations

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

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

Demonstration:

$$
\begin{aligned}
\textbf{We have } \mathrm{Var}[X] &= E[(X-E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - 2E[XE[X]] + E[X]^2 = E[X^2] - 2E[X]E[E[X]] + E[X]^2 \\
\textbf{Because } E[X+Y] &= E[X] + E[Y] \\
\textbf{Then } Var[X] &= E[X^2] - E[X]^2 \\
\textbf{Then } E[X^2] &= \mathrm{Var}[X] + E[X]^2 \\
\textbf{We have } E[f] &= f \\
\textbf{And } y &= f + \varepsilon \\
\textbf{And } E[\varepsilon] &= 0 \\
\textbf{Then } E[y] &= E[f+\varepsilon] = E[f] = f \\
\textbf{Then } \mathrm{Var}[y] &= E[(y-E[y])^2] = E[(y-f)^2] \\
\mathbf{E}[(\mathbf{y}-\hat{\mathbf{f}})^{\mathbf{2}}] &= E[y^2 + \hat{f}^2 - 2y\hat{f}] \\
&= E[y^2] + E[\hat{f}^2] - E[2y\hat{f}] \\
\textbf{Because } E[y^2] &= Var(y) + E[y]^2 \\
\textbf{And } E[\hat{f}^2] &= Var(\hat{f}) + E[\hat{f}]^2 \\
\textbf{And } E[y] &= f \\
\textbf{Then } \mathbf{E}[(\mathbf{y}-\hat{\mathbf{f}})^{\mathbf{2}}] &= \mathrm{Var}[y] + E[y]^2 + \mathrm{Var}[\hat{f}] + E[\hat{f}]^2 - 2fE[\hat{f}] \\
&= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + f^2 - 2fE[\hat{f}] + E[\hat{f}]^2 \\
\textbf{Because } f^2 - 2fE[\hat{f}] + E[\hat{f}]^2 &= (f^2 - E[\hat{f}])^2 \\
&= (E[f^2 - \hat{f}])^2 \\
\textbf{And } Var(y) = E[(y-E[y])^2] &= E[(f+\varepsilon-f)^2] \\
&= E[(\varepsilon-0)^2] = E[(\varepsilon - E[\varepsilon])^2] \\
&= Var(\varepsilon) \\
\textbf{We have finally } \mathbf{E}[(\mathbf{y}-\hat{\mathbf{f}})^{\mathbf{2}}] &= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + (f - E[\hat{f}])^2 \\
&= \mathrm{Var}[y] + \mathrm{Var}[\hat{f}] + E[(f-\hat{f})]^2 \\
&= \mathrm{Var}[\varepsilon] + \mathrm{Var}[\hat{f}] + \mathrm{Bias}[\hat{f}]^2 \\
\mathbf{E}[(\mathbf{y}-\hat{\mathbf{f}})^{\mathbf{2}}] &= \mathbf{Var}[\varepsilon] + \mathbf{Var}[\hat{\mathbf{f}}] + \mathbf{Bias}[\hat{\mathbf{f}}]^{\mathbf{2}}
\end{aligned} \tag{1}
$$