

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,  $X_1, X_2, \dots, X_p$ . We assume that there is some relationship between  $Y$  and  $X = (X_1, X_2, \dots, X_p)$ , which can be written in the very general form

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

Here  $f$  is some fixed but unknown function of  $X_1, \dots, X_p$ , and  $\varepsilon$  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

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

$$= \text{Var}(\hat{f}) + \text{Bias}(\hat{f})^2 + \text{Var}(\varepsilon) \tag{2.7}$$

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

We have

$$E[\varepsilon] = 0$$

$$E[y] = E[f + \varepsilon] = E[f] = f$$

$$\text{Var}[X] = E[X^2] - E[X]^2 \Leftrightarrow E[X^2] = \text{Var}[X] + E[X]^2$$

$$\text{Var}[y] = E[(y - E[y])^2] = E[(y - f)^2]$$

Then

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

**We have**

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

**Then**

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