2
The Bayes Error

2.1 The Bayes Problem

In this section, we define the mathematical model and introduce the notation we will use for the entire book. Let \((X, Y)\) be a pair of random variables taking their respective values from \(\mathcal{R}^d\) and \(\{0, 1\}\). The random pair \((X, Y)\) may be described in a variety of ways: for example, it is defined by the pair \((\mu, \eta)\), where \(\mu\) is the probability measure for \(X\) and \(\eta\) is the regression of \(Y\) on \(X\). More precisely, for a Borel-measurable set \(A \subseteq \mathcal{R}^d\),

\[
\mu(A) = \mathbb{P}(X \in A),
\]

and for any \(x \in \mathcal{R}^d\),

\[
\eta(x) = \mathbb{P}(Y = 1 | X = x) = \mathbb{E}(Y | X = x).
\]

Thus, \(\eta(x)\) is the conditional probability that \(Y\) is 1 given \(X = x\). To see that this suffices to describe the distribution of \((X, Y)\), observe that for any \(C \subseteq \mathcal{R}^d \times \{0, 1\}\), we have

\[
C = (C \cap (\mathcal{R}^d \times \{0\})) \cup (C \cap (\mathcal{R}^d \times \{1\})) \overset{\text{def}}{=} C_0 \times \{0\} \cup C_1 \times \{1\},
\]

and

\[
\mathbb{P}((X, Y) \in C) = \mathbb{P}(X \in C_0, Y = 0) + \mathbb{P}(X \in C_1, Y = 1) = \int_{C_0} (1 - \eta(x)) \mu(dx) + \int_{C_1} \eta(x) \mu(dx).
\]
2. The Bayes Error

As this is valid for any Borel-measurable set \( C \), the distribution of \((X, Y)\) is determined by \((\mu, \eta)\). The function \( \eta \) is sometimes called the \textit{a posteriori probability}.

Any function \( g : \mathcal{R}^d \rightarrow \{0, 1\} \) defines a \textit{classifier} or a \textit{decision function}. The error probability of \( g \) is \( L(g) = P\{g(X) \neq Y\} \). Of particular interest is the Bayes decision function

\[
g^*(x) = \begin{cases} 1 & \text{if } \eta(x) > 1/2 \\ 0 & \text{otherwise.} \end{cases}
\]

This decision function minimizes the error probability.

\textbf{Theorem 2.1.} For any decision function \( g : \mathcal{R}^d \rightarrow \{0, 1\} \),

\[ P\{g^*(X) \neq Y\} \leq P\{g(X) \neq Y\}, \]


\textit{that is, } \( g^* \) \textit{is the optimal decision.}

\textbf{Proof.} Given \( X = x \), the conditional error probability of any decision \( g \) may be expressed as

\[
P\{g(X) \neq Y|X = x\}
= 1 - P\{Y = g(X)|X = x\}
= 1 - (\eta(x)) \left(P\{Y = 1|X = x\} + (1 - \eta(x)) P\{Y = 0|X = x\}\right)
= P\{g^*(X) \neq Y|X = x\} - \left(P\{g^*(X) \neq Y|X = x\}\right),
\]

where \( I_A \) denotes the indicator of the set \( A \). Thus, for every \( x \in \mathcal{R}^d \),

\[
P\{g(X) \neq Y|X = x\} - P\{g^*(X) \neq Y|X = x\}
= \eta(x) \left(I_{g^*(x)=1} - I_{g(x)=1}\right) + (1 - \eta(x)) \left(I_{g^*(x)=0} - I_{g(x)=0}\right)
= (2\eta(x) - 1) \left(I_{g^*(x)=1} - I_{g(x)=1}\right)
\geq 0
\]

by the definition of \( g^* \). The statement now follows by integrating both sides with respect to \( \mu(dx) \). \( \Box \)

\textbf{Figure 2.1.} \textit{The Bayes decision in the example on the left is 1 if } \( x > a \), \textit{and 0 otherwise.}