THE BAYES CLASSIFIER

Consider \((X, Y)\) where

\[ X \in \mathbb{R}^d \]

\[ Y \in \{1, \ldots, K\} \]

Let \( f : \mathbb{R}^d \rightarrow \{1, \ldots, K\} \) be a classifier.

The probability of error of \( f \) is

\[ R(f) := \Pr\{f(X) \neq Y\} \]

Denote the a posteriori class probabilities by

\[ \eta_k(x) := \]

\[ k = 1, \ldots, K. \]
Theorem

The classifier

\[ f^*(x) = \arg \max_k \eta_k(x) \]

satisfies

\[ R(f^*) = \min \ R(f) \]

where the min is over all classifiers.

Terminology

- \( f^* \) is called a
- \( R(f^*) \) is called the

Proof

For convenience, assume \( X \mid Y = k \) is a continuous random variable with density \( g_k(x) \).

Let \( \pi_k := \Pr \{ Y = k \} \), the a priori class probabilities.
Consider an arbitrary classifier \( f \). Denote the decision regions

\[
\Pi_k(f) := \{ x : f(x) = k \}.
\]

Then

\[
1 - R(f) = Pr \left\{ f(x) = y \right\}
\]

To maximize this expression, we should select \( f \) such that

\[
x \in \Pi_k(f) \iff
\]

Therefore, the optimal \( f \) has

\[
f^*(x) =
\]
By Bayes rule,

\[ \eta_k(x) = \]

**Variations**

Different ways of expressing the Bayes classifier:

1. \[ f^*(x) = \arg \max_k \eta_k(x) \]

2. \[ f^*(x) = \arg \max_k \pi_k g_k(x) \]

3. When \( K = 2 \)

\[
\frac{g_2(x)}{g_1(x)} \geq \frac{\pi_1}{\pi_2}
\]

(likelihood ratio test)

4. When \( \pi_1 = \pi_2 = \ldots = \pi_K \)

\[ f^*(x) = \arg \max_k g_k(x) \]

(maximum likelihood classifier/detector)
A. \[ A_k = \sum_{k=1}^{K} \pi_k \Pr \{ f(X) = k \mid Y = k \} \]

\[ = \sum_{k=1}^{K} \pi_k \int_{\mathcal{X}} g_k(x) \, dx \]

B. \( x \in \mathcal{X}_k(x) \iff \pi_k g_k(x) \) is maximal

\[ \Rightarrow f^*(x) = \arg \max_k \pi_k g_k(x) \]

C. \( m_k(x) = \Pr \{ Y = k \mid X = x \} \)

\[ = \frac{\pi_k g_k(x)}{\sum_{l=1}^{K} \pi_l g_l(x)} \quad \xrightarrow{\text{independent of } k} \]