

# THE BAYES CLASSIFIER

---

Consider  $(X, Y)$  where

$$X \in \mathbb{R}^d$$

$$Y \in \{1, \dots, K\}$$

Let  $f: \mathbb{R}^d \rightarrow \{1, \dots, K\}$  be a classifier

The probability of error of  $f$  is

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

Denote the a posteriori class probabilities by

$$\eta_k(x) :=$$

$$k=1, \dots, K.$$

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  $X|Y=k$  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

$$\Gamma_k(f) := \{x : f(x) = k\}.$$

Then

$$1 - R(f) = \Pr \{f(X) = Y\}$$

(A)

=  
=

To maximize this expression, we should select  $f$  such that

(B)

$$x \in \Gamma_k(f) \Leftrightarrow$$

Therefore, the optimal  $f$  has

$$f^*(x) =$$

By Bayes rule,

$$\textcircled{c} \quad \eta_k(x) =$$



### Variations

Different ways of expressing the Bayes classifier:

- $f^*(x) = \arg \max_k \eta_k(x)$

- $f^*(x) = \arg \max_k \pi_k g_k(x)$

- When  $K=2$

$$\frac{g_2(x)}{g_1(x)} \begin{matrix} > \\ < \end{matrix} \frac{\pi_1}{\pi_2}$$

(likelihood ratio test)

- When  $\pi_1 = \pi_2 = \dots = \pi_k$

$$f^*(x) = \arg \max_k g_k(x)$$

(maximum likelihood classifier/detector)

Key

$$\begin{aligned} \text{A.} &= \sum_{k=1}^K \pi_k \Pr \{ f(X) = k \mid Y = k \} \\ &= \sum_{k=1}^K \pi_k \int_{\Pi_k(f)} g_k(x) dx \end{aligned}$$

$$\text{B. } x \in \Gamma_k(x) \iff \pi_k g_k(x) \text{ is maximal}$$

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

$$\text{C. } \pi_k(x) = P \{ Y = k \mid X = x \}$$

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