

## Discrimination and Classification:

### Examples:

1) Based on anthropological measurements [circumferences, volume of fossil] we want to know if a new specimen is an Ape or Human? Male or Female?

2) For each patient admitted to hospital, 19 variables are measured during first 24 hours [blood pressure, age,...] summarizing the medical symptoms which may be important indicators of patient condition. The goal is to develop a method to identify high risk patients based on measurements.

3) Based on measurements of sepal length, sepal width, petal length and petal width, decide whether a new unknown specimen is *Iris setosa*, *Iris versicolour* or *Iris virginica*.

### Formulation:

Basic goal is to assign a new object to one of the possible classes. Suppose  $C = \{1, 2, \dots, J\}$  is the set of classes and the measurements belong to a measurement space,  $\chi$ .

Purpose of Classification:

- 1) Produce an accurate classifier
- 2) Understand the predictive structure

Define:

*Measurement Space:*  $\chi$  = all possible measurement vectors.

*Set of Classes:*  $C = \{1, 2, \dots, J\}$

*Classifier:*  $\delta : \chi \rightarrow C$

Put  $A_j = \{\mathbf{x} \in \chi : \delta(\mathbf{x}) = j\}$ , then  $A_j$ 's partition  $\chi$ .

Probabilistic formulation:

Let  $P(\cdot, \cdot)$  denote a probability distribution on  $\chi \times C$ .

$P(A, j) := P(\mathbf{x} \in A, Y = j)$

*Prior class probability:*  $\pi(j) = P(Y = j)$

*Missclassification probability:*  $P[\delta(\mathbf{x}) \neq Y]$

*Bayes rule:*  $\delta_B(\cdot)$  is a classifier such that

$$P[\delta_B(\mathbf{x}) \neq Y] \leq P[\delta(\mathbf{x}) \neq Y]$$

Let  $P(A|j) = P(A, j)/\pi(j)$  be the prob. distr. of the  $j$ th class measurement vectors and assume  $P(A|j) = \int_A f_j(\mathbf{x})d\mathbf{x}$ .

Theorem:

The Bayes Rule is defined by:

$$\delta_B(\mathbf{x}) = j \iff f_j(\mathbf{x})\pi(j) = \max_i f_i(\mathbf{x})\pi(i)$$

Remark:

Bayes Rule is also ML rule: Assign  $\mathbf{x}$  to  $j$  for which  $f_j(x)\pi(j)$  is maximum.

In practice, both  $\pi(\cdot)$  and  $f_i(\cdot)$  are unknown and the Bayes rule is constructed from estimators:

A *learning sample*,  $L$ , is data  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  on  $n$  cases where  $\mathbf{x}_i \in \mathcal{X}$ ,  $y_i \in \{1, \dots, J\}$ , for  $i = 1, \dots, n$ .

Estimation of  $\pi(\cdot)$ :

Use  $\hat{\pi}(j) = (1/n)(\text{Number of } y_i \text{ in class } j \text{ in } L)$ .

Estimation of  $f_i(\cdot)$ :

Approaches:

1) parametric and 2) nonparametric

Parametric:  $f_j$ 's are known up to a few parameters

Nonparametric: estimate  $f_j$ 's directly from  $L$ .