Discrimination Analysis

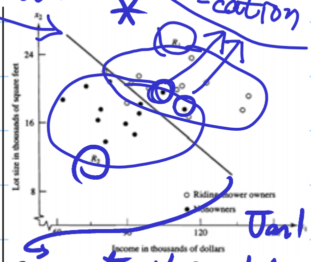
- Data and Problem: classification rule.

observed data/ training sample:

1st object →

$$\begin{array}{c|c} & \text{Var1} \quad\quad\quad\quad\quad\quad \text{Varp} \\ \underset{(n\times p)}{\mathbf{X}} = & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \end{array} \quad \begin{array}{c} \textbf{category} \\ 1 \\ 1 \\ \vdots \\ k \end{array}$$

k = # of different categories



seperate the whole data space into two subsets R_1 & R_2

future observation/test sample: $(x_{f1} \quad x_{f2} \quad \ldots \quad x_{fp})$ **?**

- ➢ objective: use certain *observed* measurements **X** of some objects whose categories or grouping are *known*, to *determine a rule* that can be used to assign a new object (whose category is *unknown*) to one of the pre-specified categories
- ➢ discrimination analysis also known as *pattern recognition*, (*statistical*) *classification*, or *numerical taxonomy*
- ➢ examples for k=2

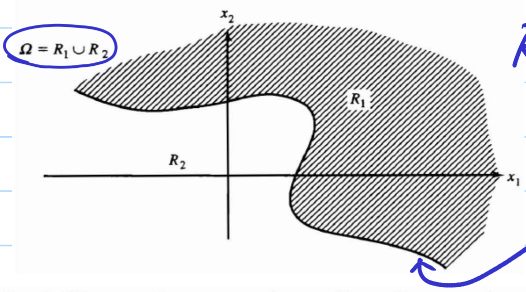
categories classes -

Populations π_1 and π_2	Measured variables **X**
5. Purchasers of a new product and laggards (those "slow" to purchase).	Education, income, family size, amount of previous brand switching.
6. Successful or unsuccessful (fail to graduate) college students.	Entrance examination scores, high school grade-point average, number of high school activities.
9. Alcoholics and nonalcoholics.	Activity of monoamine oxidase enzyme, activity of adenylate cyclase enzyme.

- ➢ Q: Why the categories of some objects are known, some unknown? Here are some possible conditions:
  - incomplete knowledge of "future" performance
  - "perfect" information required destroying the object
  - unavailable or expensive information
- ➢ a good classification procedure should
  - result in few misclassifications
  - take "prior probabilities of occurrence" into account

    If n objects are randomly sampled, the # of objects in ith category should reflect it.
    - example. There tend to be more financially sound firms than bankrupt firm. If we really believe that the prior probability of a financially distressed and ultimately bankrupted firm is very small, then one should classify a randomly selected firm as non-bankrupt unless the data overwhelmingly favor bankruptcy.
  - consider the cost
    - Suppose that classifying a π_1 object as belonging to π_2 (type I error) represents a more serious error than classifying a π_2 object as belonging to π_1 (type II error). Then, one should be cautious about making the former assignment
    - example. Diagnosis of a potentially fatal illness

- Separation and Classification for Two Populations
  - Modeling of the data and problem
    - Let $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ be the probability density functions associated with the $p \times 1$ vector random variable $\mathbf{X}$ for the populations π_1 and π_2, respectively.
    - Let Ω be the sample space—that is, the collection of all possible observations $\mathbf{x}$.
    - Let R_1 be that set of $\mathbf{x}$ values for which we classify objects as π_1 and $R_2 = \Omega - R_1$ the sets R_1 and R_2 are mutually exclusive and exhaustive.

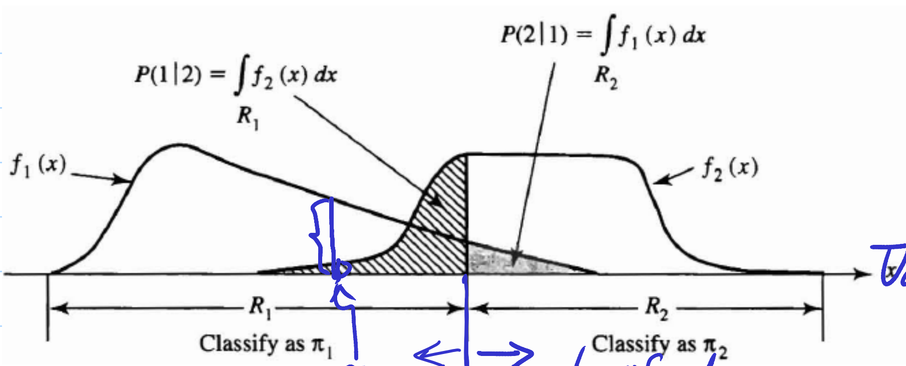


$R_1 \cap R_2 = \phi$

decision rule.

    - Let p_1 be the *prior* probability of π_1 and p_2 be the *prior* probability of π_2, where $p_1 + p_2 = 1$.
    - The costs of misclassification can be defined by a cost matrix:

		Classify as: π_1	Classify as: π_2
True population:	π_1	0	$c(2\mid 1)$
	π_2	$c(1\mid 2)$	0



Var1 ∈ R

x_0

classification rule.

decision

  - Calculation of some probabilities
    - The conditional probability, $P(2\mid 1)$, of classifying an object as π_2 when, in fact, it is from π_1 is

$$P(2\mid 1) = P(\mathbf{X} \in R_2 \mid \pi_1) = \int_{R_2 = \Omega - R_1} f_1(\mathbf{x})\,d\mathbf{x}$$

    - the conditional probability, $P(1\mid 2)$, of classifying an object as π_1 when it is really from π_2 is

$$P(1\mid 2) = P(\mathbf{X} \in R_1 \mid \pi_2) = \int_{R_1} f_2(\mathbf{x})\,d\mathbf{x}$$

    - the overall probabilities of correctly or incorrectly classifying objects are
      - $P(\text{observation is correctly classified as } \pi_1) = P(\text{observation comes from } \pi_1 \text{ and is correctly classified as } \pi_1)$

$$\boxed{\int_{R_1} f_1(x)dx} \quad = P(\mathbf{X} \in R_1 \mid \pi_1)P(\pi_1) = P(1\mid 1)\,p_1$$

- P(observation is misclassified as π_1) = P(observation comes from π_2 and is misclassified as π_1) $= P(\mathbf{X} \in R_1 | \pi_2)P(\pi_2) = P(1|2)p_2$
- P(observation is correctly classified as π_2) = P(observation comes from π_2 and is correctly classified as π_2) $= P(\mathbf{X} \in R_2 | \pi_2)P(\pi_2) = P(2|2)p_2$
- P(observation is misclassified as π_2) = P(observation comes from π_1 and is misclassified as π_2) $= P(\mathbf{X} \in R_2 | \pi_1)P(\pi_1) = P(2|1)p_1$

➢ *expected cost of misclassification* (ECM) criterion

$$E(\text{cost}) = \text{ECM} = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

⇒ a reasonable classification rule should have ECM as small as possible

➢ **Result 11.1.** The regions R_1 and R_2 that minimize the ECM are defined by the values $\mathbf{x}$ for which the following inequalities hold:

decision Thm
Most powerful test

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\left(\begin{array}{c}\text{density}\\ \text{ratio}\end{array}\right) \geq \left(\begin{array}{c}\text{cost}\\ \text{ratio}\end{array}\right)\left(\begin{array}{c}\text{prior}\\ \text{probability}\\ \text{ratio}\end{array}\right)$$

$$\text{ECM} = c(2|1)\overbrace{\int_{R_2} f_1(x)\,dx}^{=1-\int_{R_1} f_1\,dx} \cdot P_1 + c(1|2)\int_{R_1} f_2(x)\,dx \cdot P_2$$

$$= \int_{R_1} \underbrace{-c(2|1)P_1 f_1(x) + c(1|2)P_2 f_2(x)}_{<0}\,dx$$

$$R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\left(\begin{array}{c}\text{density}\\ \text{ratio}\end{array}\right) < \left(\begin{array}{c}\text{cost}\\ \text{ratio}\end{array}\right)\left(\begin{array}{c}\text{prior}\\ \text{probability}\\ \text{ratio}\end{array}\right)$$

$R_1 \cup R_2 = \Omega$

$R_1 \cap R_2 = \phi$

- Note. Implementation of the minimum ECM rule requires only three *ratios*
- special cases of minimum expected cost regions

When we do not have information about what the ratio should be, we often assume it.

  - $p_2/p_1 = 1$ (equal prior probabilities)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1|2)}{c(2|1)} \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1|2)}{c(2|1)}$$

  - $c(1|2)/c(2|1) = 1$ (equal misclassification costs)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1}$$

  - $p_2/p_1 = c(1|2)/c(2|1) = 1$ or $p_2/p_1 = 1/(c(1|2)/c(2|1))$ (equal prior probabilities and equal misclassification costs)

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \qquad R_2: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1$$