Pattern Recognition

Introduction

- Pattern Recognition (PR) is an important aspect of Artificial Intelligence (AI)

What is AI ?

- AI is the part of computer science concerned with designing intelligent computer systems that exhibit the characteristics we associate with intelligence in human behavior - understanding language, learning, reasoning and solving problems etc.

Some topics of AI

- *Natural Language Processing* to enable it to communicate successfully in natural languages.
- *Knowledge Representation* to store what it knows.
- *Automated Reasoning* to use stored information to answer questions and to draw new conclusions.

Contd…

- *Machine Learning* to adapt to new circumstances and to detect patterns.
- *Computer Vision* to perceive objects.
- *Robotics* to manipulate objects and move about.

- **Note** : In every aspect of human intelligence recognition and / or learning is involved.

Some Examples of Learning…

- When we see we recognize or learn
- When we hear we recognize or learn
- When we touch we recognize or learn
- When we smell we recognize or learn
- When we taste we recognize or learn

Contd…

- In fact, whatever we do with our five sense organs is either a kind of recognition or learning.
- Thus recognition is an important part of human intelligence. And accordingly, pattern recognition has become an important part of Artificial Intelligence.
- The problem of pattern recognition can be divided into two parts :

Part-1…

- *First part* : It is concerned with the study of recognition mechanism of patterns by human and other living organism. This part is related to the disciplines like physiology, psychology, biology, etc.

Part-2…

- *Second part* : It deals with the development of theory and techniques for designing a device which can perform these recognition tasks automatically. This area is related to engineering, computer and information sciences. In this curriculum, we shall be dealing with the second part i.e. the problems of automatic machine recognition of patterns.

Role of memory in pattern recognition

- There are two aspects of the memory in this process.
- First, the part which holds the information we can recall, e.g., a poem, a face, a vocabulary, a theorem etc. To use computer terminology, this form of memory is addressable, and its content can be recalled In this type of problem, PR is a two-step process :
- Knowledge representation
- Searching

Contd…

- Secondly, the information which presumably is stored somewhere, but which we cannot retrieve, such as we cannot describe how we balance when we walk, how we recognize a speech or how we drive a car or similar aspects of pattern processing, although the information must be stored somewhere in our brain and nervous system.

Contd…

- This form of memory is not addressable, and its content can not be recalled. In this case, the PR process is a two-fold task :
  - Developing some decision rules based on previous knowledge (**Learning**)
  - Use it for taking decision regarding an unknown pattern (**Classification**)

Some of the important Pattern Recognition application areas

1) **MAN-MACHINE Communication**-

(a) Automatic Speech Recondite,

(b) Speaker Identification,

(c) OCR Systems,

(d) Cursive Script Recognition

(e) Image Understanding.

Contd…

2) **BIOMEDICAL Applications-**

(a) ECG, EEG, EMG Analysis,

(b) Cytological, Histological and other Stereological Applications,

(c) X-ray Analysis

(d) Medical Diagnostics.

Contd…

3) **APPLICATION IN PHYSICS**–

(a) High Energy Physics,

(b) Bubble Chamber and other Forms of Track Analysis.

4) **CRIME AND CRIMINAL DETECTION-**

Contd…

(a) Fingerprint,

(b) Hand Writing

(c) Speech Sound and

(d) Photographs

5) **NATURAL RESOURCES STUDY AND Estimation-**

Contd…

(a) Agriculture,
(b) Hydrology,
(c) Forestry,
(d) Geology,
(e) Environment,
(f) Cloud Pattern,
(g) Urban Quality

Contd…

6) **STEREOLOGICAL APPLICATIONS-**

a) Metal Processing.

b) Mineral Processing and

c) Biology.

7) **MILITARY APPPICATIONS-**

Contd…

(a) Detection of Nuclear Explosions,
(b) Missile Guidance and Detection
(c) Radar and Sonar Signal Detection,
(d) Target Identification.
(e) Naval Submarine Detection etc.

What is Pattern ?

- A pattern is a description of an object that we are going to recognize.
- Description means a set of measurements
- But all these measurements are not significant in the context of recognition

Example

- Given a set of professional basketball players and wrestlers, how to recognize an individual of the set to be a player of any one of these two games.

- Measurements : Age, Qualification, Height, Weight
- Basis of feature selection : Basketball players are taller and slimmer, whereas wrestlers are comparatively shorter and heavier.
- Feature: Height, Weight.

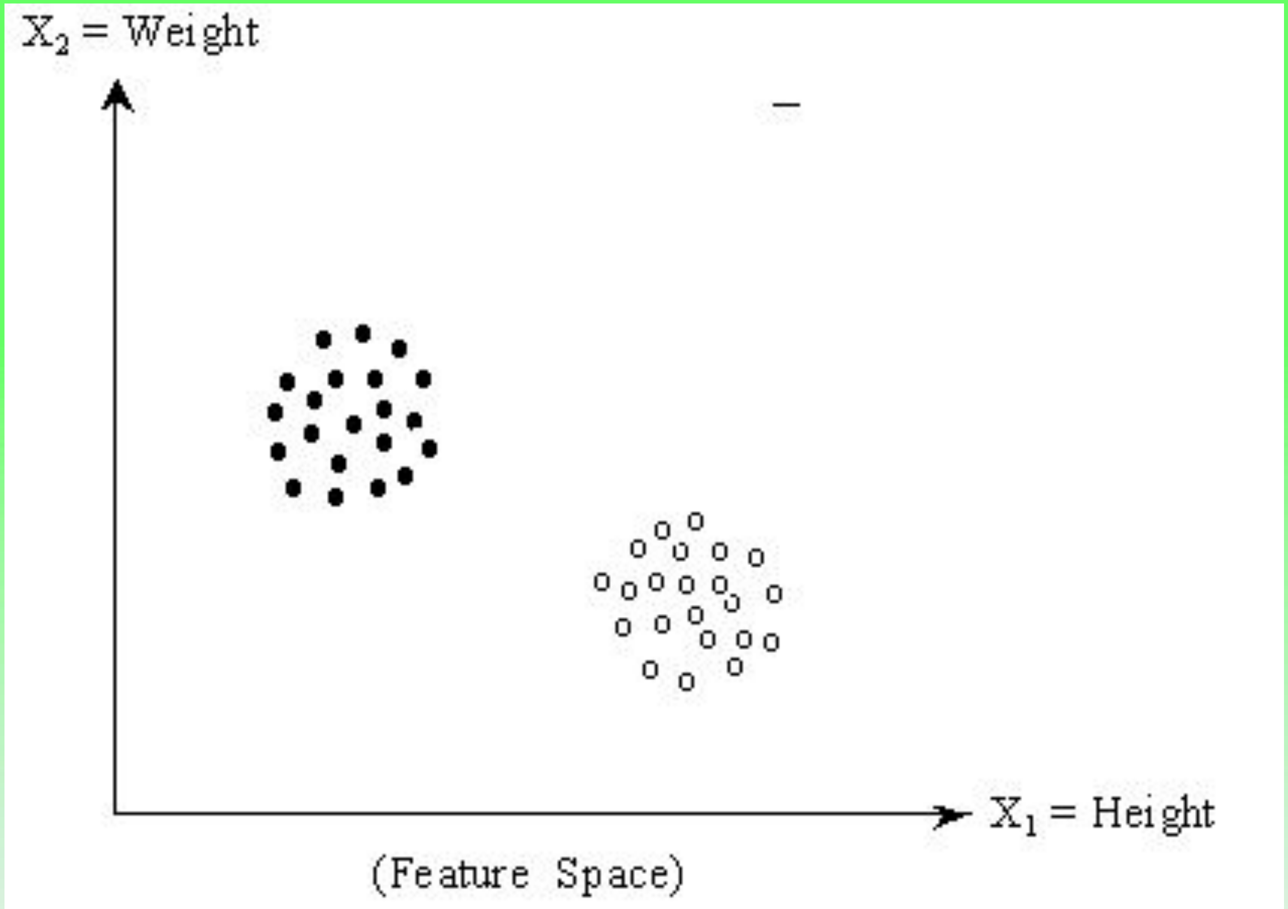
Pattern vector or Feature vector

- When each feature is considered as a component of a vector, it is called a Pattern vector or Feature vector.
- **Feature Space**

A pattern vector is represented as a point in *n*-dimensional Euclidian space

Contd…

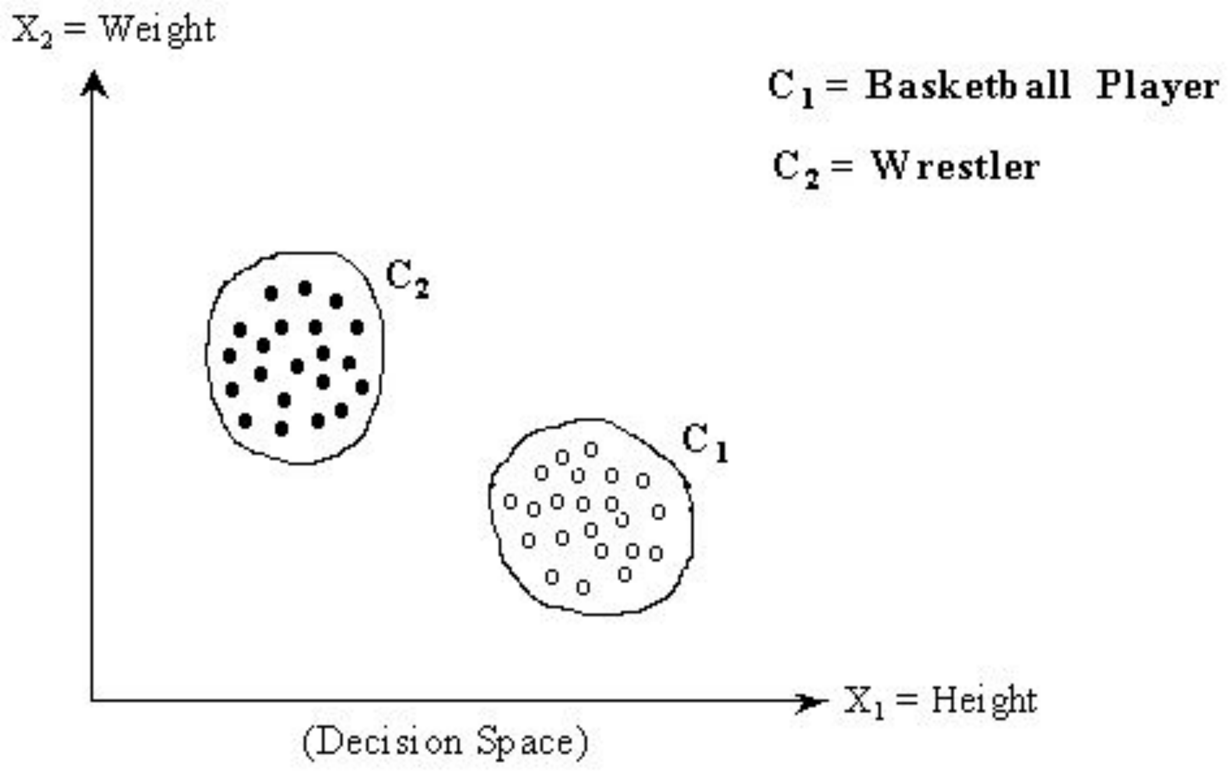
-



Decision Space

- When different regions of the feature space are identified to belong to different classes, then such a space is called a decision space

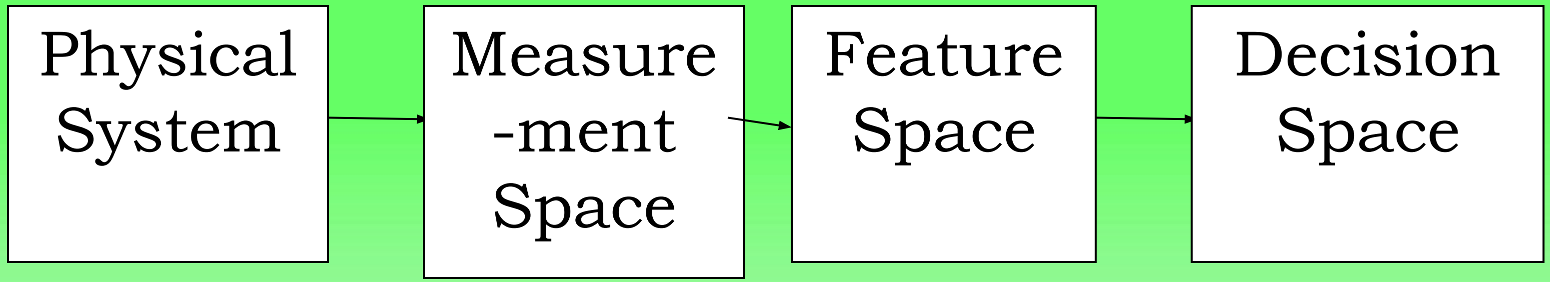
Contd…



X_2 = Weight
C_1 = Basketball Player
C_2 = Wrestler
C_2
C_1
X_1 = Height
(Decision Space)

Operating stages in a Pattern Classifier

Steps in a typical PR System



Classification of PR

- Pattern Recognition
  - Supervised
  - Unsupervised
    - Non-Hierarchical (Partitional)
    - Hierarchical

Different approaches in PR techniques/methodologies

- Pattern Classification by Decision Functions
- Pattern Classification by Distance Functions
- Pattern Classification by Likelihood Functions
- Trainable Pattern Classifiers-The Deterministic Approach
- Trainable Pattern Classifiers-The Statistical Approach
- Syntactic Pattern Recognition
- Pattern Preprocessing and Feature Selection

Pattern Classification by Decision Functions

- Let $C_1, C_2, \ldots C_j, \ldots C_m$ be designated as the *m* possible pattern classes in an *N*-dimensional feature vector space Ω_X and let

$$X = [x_1, x_2, \ldots, x_n, \ldots, x_N]'$$

Contd…

be the unknown pattern vector, where x_n represents the *nth* feature measurement.

- If the pattern *X* is a member of class C_k the decision function $D_k(X)$ associated with the class C_k,k=1.2,…m must then possess the largest value. In other words,

Contd…

Decide X€C_k if $D_k(X) > D_j(X)$ with $K \neq j$, $k,j=1,2,\ldots,m$.

- Ties are resolved arbitrarily. The decision boundary in the *N*-dimensional feature space Ω_X, between regions associated with classes C_k and C_j respectively would be governed by the expression

Contd…

$D_k(X) - D_j(X) = 0$ with $k \neq j$, $k,j=1,2,\ldots,m$.

- Let us now consider two hypothetical pattern classes in R^2 as shown below

Contd…

-

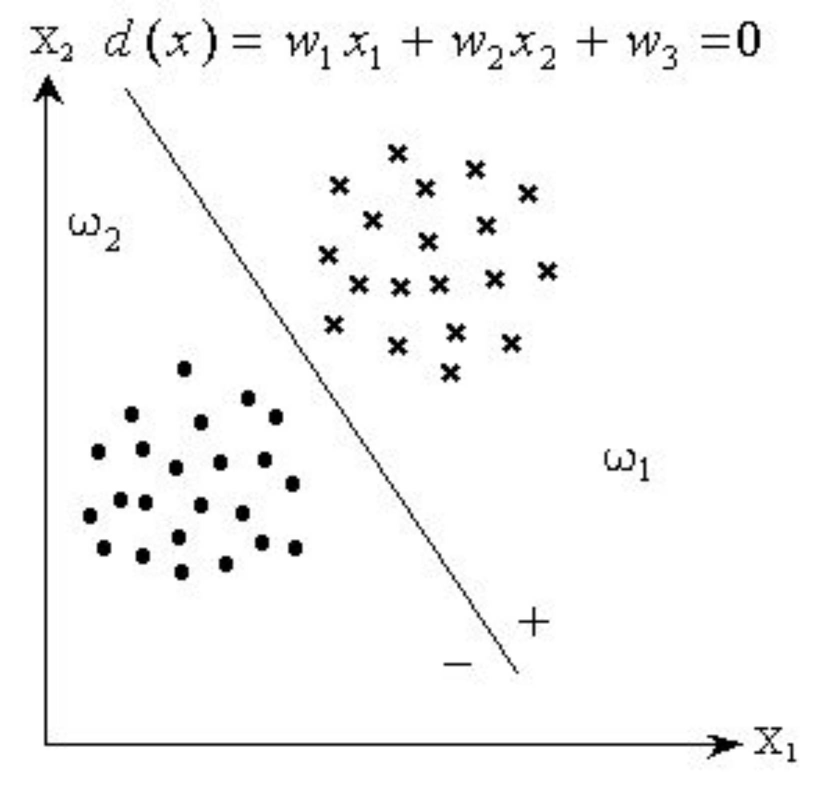


Figure : A simple decision function for two pattern classes

Contd…

- Let $d(x)=w_1x_1+w_2x_2+w_3$ be the equation of a separating line where the w's are parameters and $x_1,x_2$are the general coordinate variables.
- It is clear from the figure that

$X€C_1$(or ω_1), if $D(X)>0$ and $X €C_2$(or ω_2) if $D(X)<0$

Contd…

- The success of the said pattern classification scheme depends on two factors :
  - The form of d(X)
  - One’s ability to determine its coefficients
- The first problem is directly related to the geometrical properties of the pattern classes under consideration. Unless some *a priori* information is available,

Contd…

the only way to establish the effectiveness of a chosen decision function is by direct trial.

- Once a certain function (or functions if more than two classes are involved) has been selected, the problem becomes the determination of the coefficients. Several adaptive and training schemes exist that can solve this problem.

Contd…

Sometimes, some sample pattern can be utilized in order to determine the coefficients which characterize an already specified decision function.

General *n*-dimensional form of decision function

- A general linear decision function is of the form

$d(X)=w_1x_1+w_2x_2+\ldots+w_nx_n+w_{n+1}$

$= W'_0\, X+w_{n+1}$ where

$W_0=(w_1,w_2,\ldots,w_n)'$

This vector is referred to as *weight* or *parameter* vector.

Contd…

- It is a widely accepted convention to append a 1 after the last component of *all* pattern vectors

  $d(X)=W'X$

  where $X=(x_1,x_2.....x_n,1)'$ and $W=(w_1,w_2,......w_n,w_{n+1})'$ are called *augmented* pattern and weight vectors, respectively. Since the same quantity is equally appended to all patterns, the basic geometrical properties of the pattern classes are not disturbed.

Contd…

- In the two-class case a decision function d(X) is assumed to have the property

$$d(X) = W'X \begin{cases} > 0 \quad if \quad X \in \omega_1 \\ \hline < 0 \quad if \quad X \in \omega_2 \end{cases}$$

When we have more than two classes, denoted by $w_1, w_2 \ldots w_M$ we consider the following multi-class cases.

Contd…

- *Case 1*. Each pattern class is separable from the other classes by a single decision surface. In this case there are *M* decision functions with the property

$$d_i(X) = W_i' X = \begin{cases} > 0 \; if \;\; X \in \omega_i \\ \hline < 0 \; otherwise \end{cases}, i = 1,2,\ldots,M$$

Contd…

Where $W_i=(w_{i1},w_{i2},\ldots.,w_{in},w_{i,n+1})'$ is the weight vector associated with the i^{th} decision function

- *Case 2.* Each pattern class is separable from every other individual class by a distinct decision surface, that is the classes are pairwise separable. In this case there are M(M-1)/2 (the combination of *M* classes taken two at a time) decision surface. The decision functions here are of the from $d_{ij}(X)=W_{ij}'X$ and have the property that,

Contd…

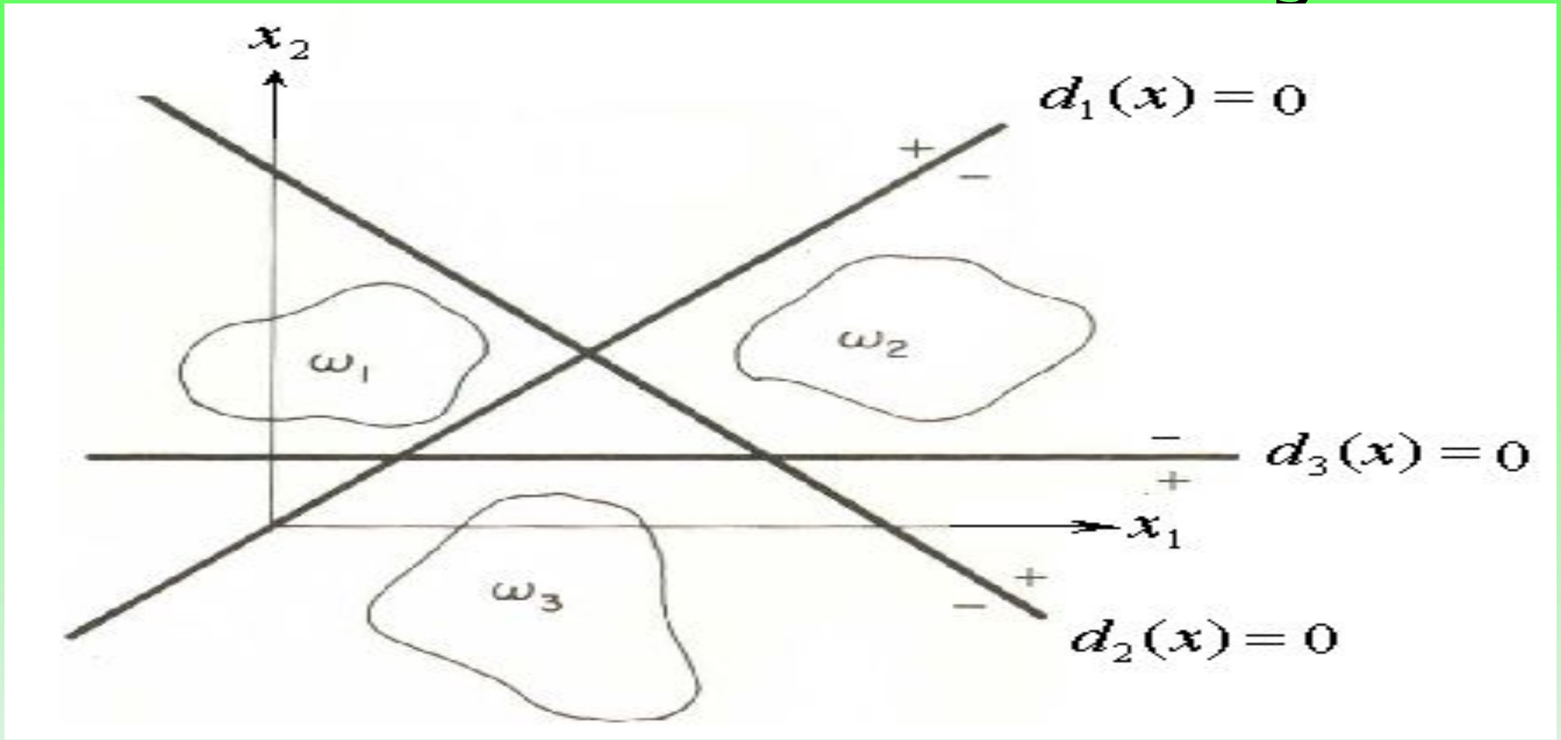
if x belongs to class ω_i then $d_{ij}(X)>0$ for all $j\neq i$

These functions also have the property that $d_{ij}(X)=-d_{ji}(X)$

It is not uncommon to find problems involving a combination of class 1 and 2. These situations require fewer than the M(M-1)/2 decision surfaces which would be needed if all the classes were only pairwise separable.

Example-1

- A simple example of multi-class case-1 is shown in the following



Contd…

- It is noted that each class is separable from the rest by a single decision boundary.
- if x belongs to class ω_1, then $d_1(X)>0$ while $d_2(X)<0$ and $d_3(X)<0$
- The boundary between class ω_1 and the other classes is given by the values of x for which $d_1(X)=0$
- As a numerical illustration assume that the decision functions of the above Figure to be

$$d_1(X) = -x_1 + x_2, \quad d_2(X) = x_1 + x_2 - 5, \quad d_3(X) = -x_2 + 1$$

Contd…

- The three decision boundaries are, therefore,

$-x_1+x_2=0$, $x_1+x_2-5=0$, $-x_2+1=0$

- For example, suppose that it is desired to classify the pattern $x=(6,5)'$. Substituting this pattern into the three decision function yields

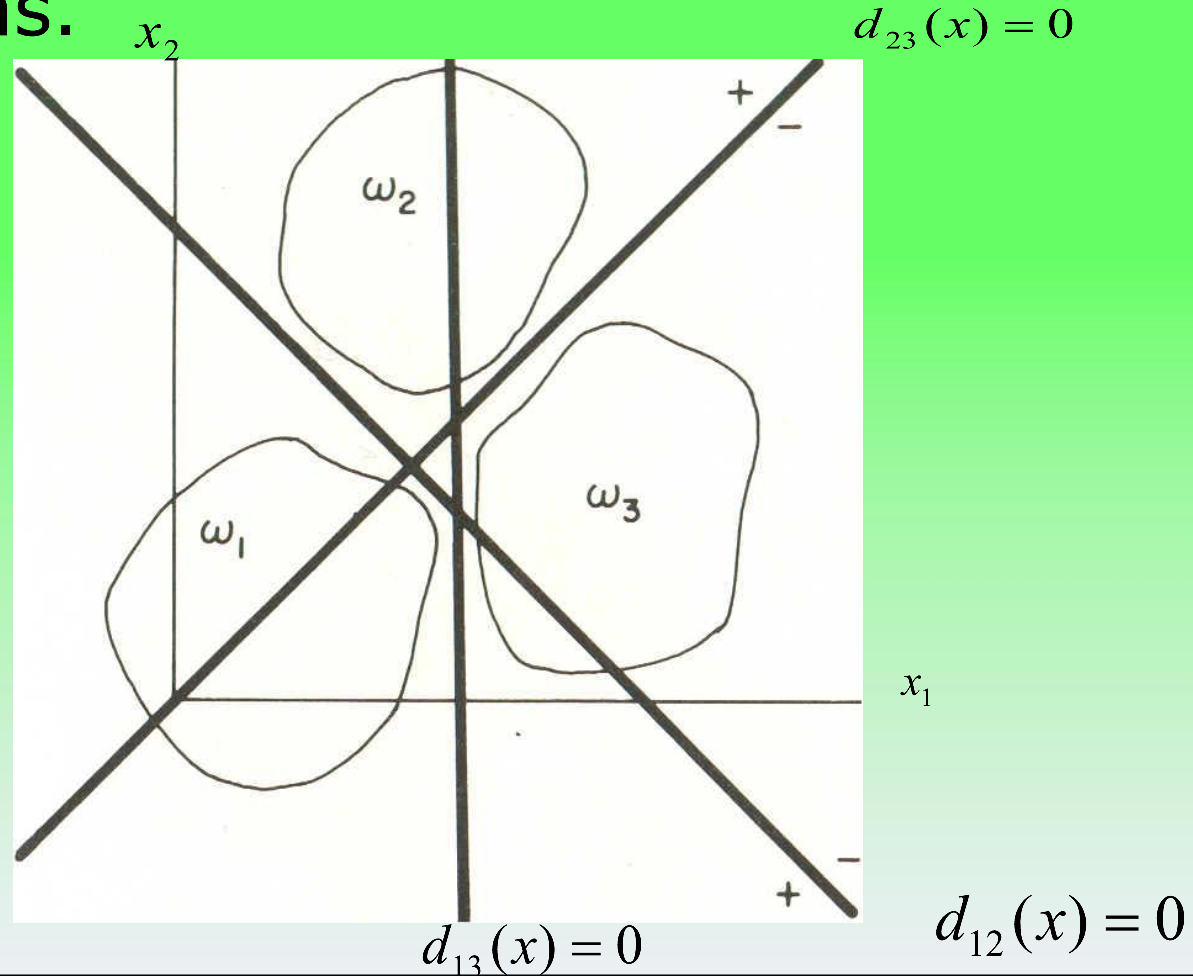
Contd…

$d_1(X)=-1, \quad d_2(X)=6, \quad d_3(X)=-4$

- Since $d_2(X)>0$ while $d_1(X)<0$ and $d_3(X)<0$ the pattern is assigned to class ω_2

Example2

- The following Figure illustrates three pattern classes separable under case 2 conditions.



Contd…

- Here no class is separable from the other by a single decision surface. Each boundary shown is capable of separating just two classes.
- Let us assume the following numerical values:

$d_{12}(X) = -x_1 - x_2 + 5$, $d_{13}(X) = -x_1 + 3$, $d_{23}(X) = -x_1 + x_2$

Contd…

- The decision boundaries are again determined by setting the decision function equal to zero. The decision regions, however, are now given by the positive sides of multiple decision boundaries.
- For example, if x belongs to class ω_i then $d_{12}(X)>0$ and $d_{13}(X)>0$

Contd…

- The value of $d_{23}(X)$ in this region is irrelevant since $d_{23}(X)$ is not related to class ω_1.
- Suppose that it is desired to classify the pattern $x=(4,3)'$. Substitution of this pattern into above decision functions yields

$$d_{12}(X)=-2,\quad d_{13}(X)=-1,\quad d_{23}(X)=-1$$

Contd…

- Since these functions have the property that $d_{ij}(X)=-d_{ji}$ (X).
- It follows that $d_{21}(X)=2$, $d_{31}(X)=1$, $d_{32}(X)=1$
- That is, $d_{3j}(X)>0$ for $j=1,2$
- Therefore we assign the pattern to class ω_3.

Geometrical properties of linear decision functions (Hyperplane Properties)

- In the two-class problem, as well as multi-class cases 1 and 2, the equation of the surface separating the pattern classes is obtained by letting the decision functions be equal to zero. In other words, in the two-class case the surface between the two pattern populations is given by the equation

Contd…

$$d(X)=w_1x_1+w_2x_2+\ldots+w_nx_n+w_{n+1}=0 \ldots\ldots\ldots\ldots(1)$$

- In case 1, the equation of the boundary ω_i between and the remaining classes is given by

$$d_i(X)=w_{i1}x_1+w_{i2}x_2+\ldots.+w_{in}x_n+w_{i,n+1}=0 \ldots\ldots\ldots(2)$$

Contd…

- Similarly, In case 2, the boundary between ωi and ωj is given by

$$d_{ij}(X)=w_{ij1}x_1+w_{ij2}x_2+....+w_{ijn}x_n+w_{ij,n+1}=0$$

- In general, the equation of the decision surface between classes ω_i and ω_j is given by

$$d(X)=w_1x_1+w_2x_2+...+w_nx_n+w_{n+1}=0$$

Contd…

$$= W'_0X+w_{n+1}=0 \ldots\ldots\ldots\ldots.(3)$$

Where $W_0=(w_1,w_2\ldots..w_n)'$

Equation (3) is recognized as the equation of a line when n=2 and as the equation of a plane when n=3. When n>3, Eq. (3) is the equation of a *hyperplane*.

Contd…

- A "hyperplane" is schematically shown in the following figure.

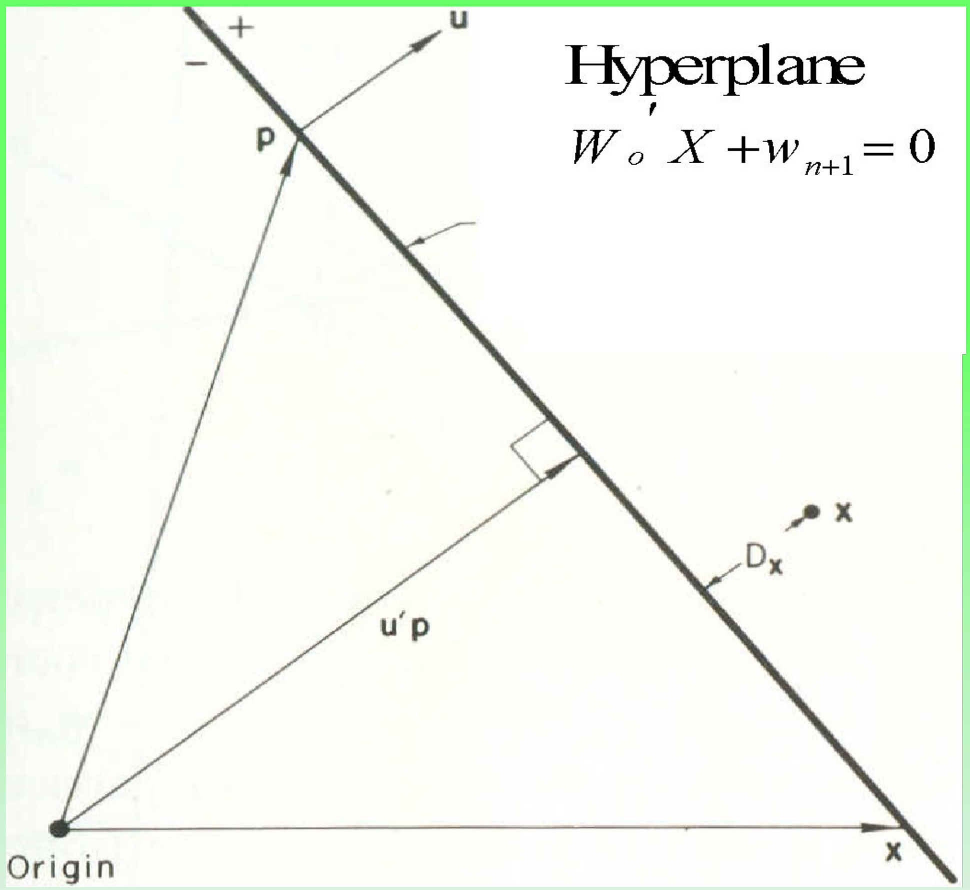


Figure 1. Some geometrical properties of hyperplanes

Contd…

- Let **u** be a unit normal to the hyperplane at same point **p** and oriented to the positive side of the hyperplane. From geometrical considerations the equation of the hyperplane may be written as

  $u'(x-p)=0$……………(4)

  or

  $u'x=u'p$………………(5)

Contd…

- Dividing Eq. (3) by $\|W_o\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ results in the equation

$$\frac{W_o' X}{\|W_o\|} = -\frac{w_{n+1}}{\|W_o\|} \qquad \text{..........................(6)}$$

Comparing Eqs. (5) and (6), we see that the unit normal to the hyperplane is given by

Contd…

$$u = \frac{W_o}{\|W_o\|} \qquad \text{..................(7)}$$

Also,

$$u'p = -\frac{w_{n-1}}{\|W_o\|} \qquad \text{..................(8)}$$

Contd…

- It is seen by comparing Fig. 1 and Eq. (8) that the absolute value of u'p represents the normal distance from the origin to the hyperplane. Denoting this distance by $D_{u'}$ we obtain

$$D_u = \frac{|w_{n+1}|}{\|W_o\|} \qquad \text{......................(9)}$$

Contd…

- From Fig. 1 it also reveals that the normal distance D_x from the hyperplane to an arbitrary point x is given by

$$D_x = |u'x - u'p|$$

$$= \left| \frac{W_o{}' X}{\|W_o\|} + \frac{w_{n+1}}{\|W_o\|} \right| \qquad = \left| \frac{W_o{}' X + w_{n+1}}{\|W_o\|} \right|$$

Contd…

- The unit normal **u** indicates the orientation of the hyperplane. If any component of **u** is zero, the hyperplane is parallel to the coordinate axis which corresponds to that component. Therefore, since, $u=W_o/||W_o||$ it is possible to tell by inspection of the vector W_0 weather a particular hyperpalne is parallel to any of the coordinate axes. We also see from Eq. (9) that if $w_{n+1}=0$ the hyperpalne passes through the origin.

Generalized Decision Functions

- A generalized form of the decision function can be defined as :

$$d(X) = w_1 f_1(X) + w_2 f_2(X) + \cdots + w_K f_K(X) + w_{K+1}$$
$$= \sum_{i=1}^{K+1} w_i f_i(X) \qquad \text{.............................(1)}$$

Contd…

- where the $\{f_i(X)$,i=1,2,…,K, are real, single-valued functions of the pattern X,$f_{k+1}(X)=1$, and K+1 is the number of terms used in the expansion. The Eq.(1) represents an infinite variety of decision functions, depending on the choice of the functions $\{f_i(X)\}$ and on the number of terms used in the expansion.

Contd…

- In spite of the fact that the above equation could represent very complex decision functions, it is possible to treat these functions as *linear* by virtue of a transformation. For that we define a vector X* whose components are the functions $f_i(X)$, that is,

Contd…

$$X^* = \begin{pmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_K(X) \\ 1 \end{pmatrix} \quad \text{…………………………………….(2)}$$

Using Eq. (2), we may express (1) as $d(X) = W'X^*$ …………(3)

Contd…

where $W=(w_1,w_2,\ldots,w_k,w_{k+1})'$.

- Once evaluated, the functions $\{f_i(X)\}$ are nothing more than a set of numerical values, and X* is simply a *K*-dimensional vector which has been augmented by 1. Therefore, Eq. (3) represents a linear function with respect to the new patterns X*. Thus any decision function of the form shown in Eq. (1) can be treated as linear by virtue.

Thank You