PATTERN REPRESENTATION

Dr. Umarani Jayaraman
Assistant Professor

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING,
KANCHEEPURAM

Data Structures for Pattern Representation

- 1. Vector representation
- 2. String representation
- 3. Logical representation/description
- 4. Fuzzy/rough representation
- 5. Tree/ Graph representation

1. Vector representation

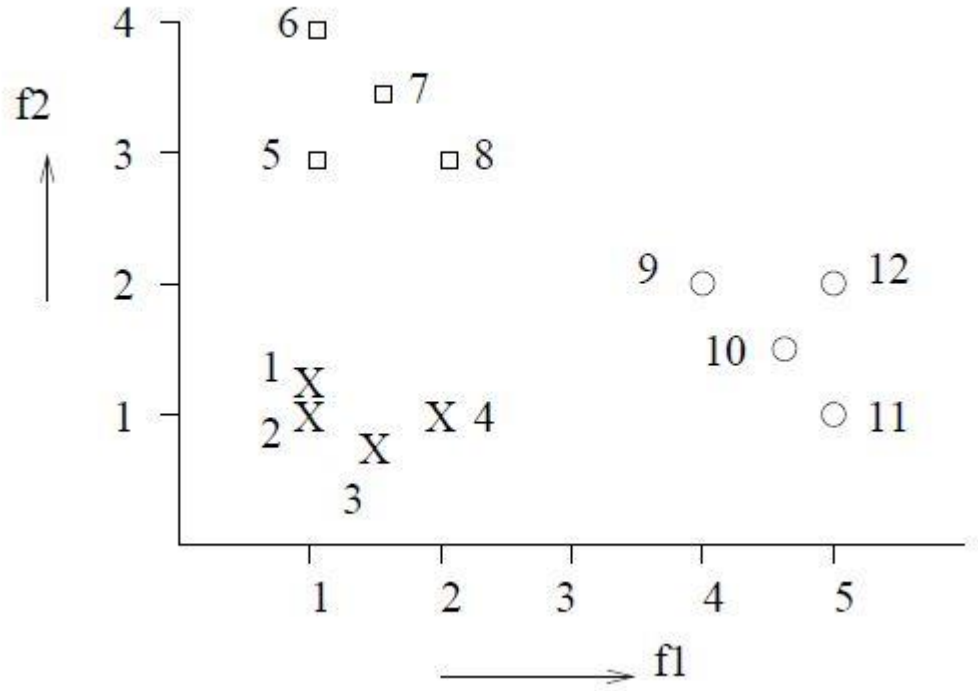
- Patterns as vectors:
  - An obvious representation of pattern is a vector. Each element of the vector represent an attribute of the pattern.
  - e.g. : Consider a spherical object (x_1,x_2,y) where x_1 denotes weight and x_2 denotes diameter and y is the class object.
  - Using vector representation , a set of patterns can be represented as :

Vector representation (contd)

Pattern 1 : (1,1.25,1)
Pattern 2 : (1,1,1)
Pattern 3 : (1.5,0.75,1)
Pattern 4 : (2,1,1)
Pattern 5 : (1,3,2)
Pattern 6 : (1,4,2)
Pattern 7 : (1.5,3.5,2)
Pattern 8 : (2,3,2)
Pattern 9 : (4,2,3)
Pattern 10 : (4.5,1.5,3)
Pattern 11 : (5,1,3)
Pattern 12 : (5,2,3)

Vector representation (contd)

- The above mentioned patterns is plotted as follows : (f1 and f2 are the feature axes)



2. Patterns as Strings

- The string may be viewed as a sentence in a language.
- For e.g. A gene can be defined as a region of the chromosomal DNA constructed with four nitrogenous bases.

  A-Adenine, G-Guanine, C-Cytosine and T-Thymine

  GAAGTCCAG----------

3. Patterns as Logical Descriptors

- Patterns can be represented as a logical description of the form:

  $(x_1 = a_1 \text{ or } a_2)$ AND $(x_2 = b_1 \text{ or } b_2)$ AND $(x_3 = c_1)$ ….

- e.g. (color = red or white) AND (make = leather) AND (shape = sphere)
- It represents a cricket ball.

4. Fuzzy and Rough Patterns

- Fuzziness is used where it is not possible to make precise statements.
- For e.g. , "If x_1 is small and x_2 is large, then class 3"

  <small, large, 3>
- It is therefore used to model subjective, incomplete and imprecise data.
- In a fuzzy set, the objects belong to a set depending on a membership value which varies from 0 to 1.

Fuzzy and Rough Patterns (contd)

- It can also be used in cases where there are uncertain or missing values : X=(?,6.2,7) or Y=([0,1],6.2,7) with no missing values.

Fuzzy and Rough Patterns (contd)

- The values of the features may be rough values. Such feature values are called rough patterns.
- A rough value consist of an upper and a lower bound.
- For e.g. power 'P' can be represented as:

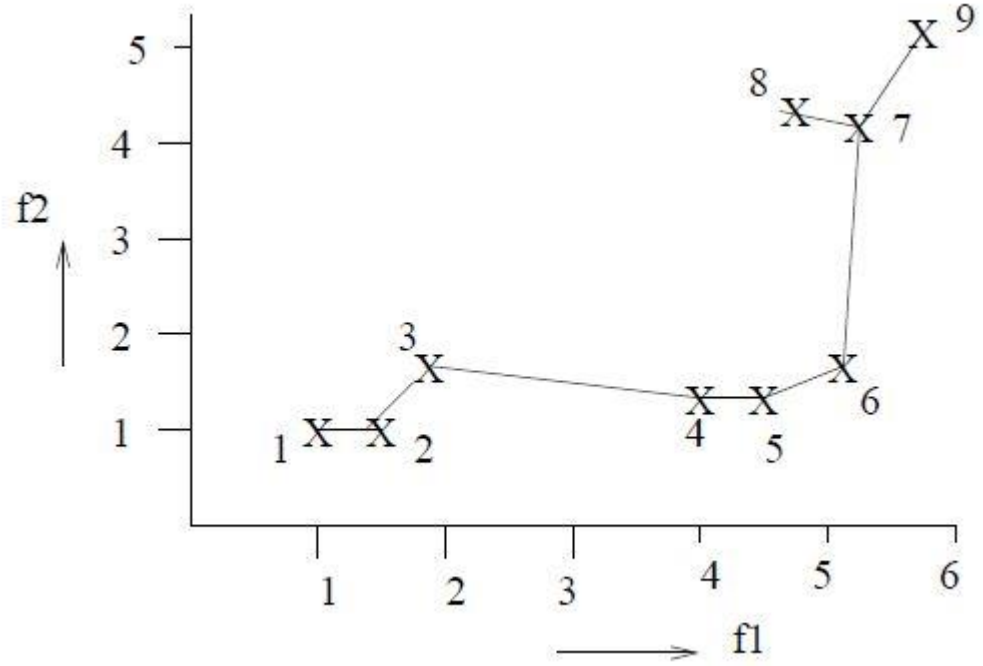
  < 230, 5.2, (50,49,51)>

  <Voltage, Current, Frequency>

5. Patterns as Trees and Graphs

- Each node in the tree or graph may represent one or more patterns.
- For e.g. , Minimum Spanning Tree (MST), K-D Tree, Frequent Pattern Tree (FP Tree), Delauney Tree (DT), R-Tree.
- The complete set of patterns can be represented using the minimum spanning tree as follows:

Patterns as Trees and Graphs (Contd)



Example: Clustering using MST

- The above figure shows a pattern set of 9 points. The minimum spanning tree is shown for the 9 points.
- MST can be used for clustering.
- Find the edge with the maximum distance in the MST and delete that edge. Here 3-4 would be deleted.
- Then find the edge with the maximum distance in the resulting tree and delete it. Now 6-7 would be deleted.
- If we stop deleting edges at this point, we get three separate components. Each of these would be one cluster.
- If we want four clusters, we can again find the edge with the maximum distance and delete it.
- An important property is that each pattern is a numerical vector and an edge is characterized by the distance between two such vectors.

Summary

- 1. Vector representation
- 2. String representation
- 3. Logical representation/description
- 4. Fuzzy/rough representation
- 5. Tree/ Graph representation

THANK YOU