INTRODUCTION

- This course deals with pattern recognition. A **pattern** is either a physical object, for example a *book or a chair* or an abstract notion, like *style of talking*, or *style of writing*. It is also a shared property of a set of objects; for example, *chairs*, *rectangles*, or *blue* colored objects. We illustrate using ellipses and rectangles shown in Figure 1.

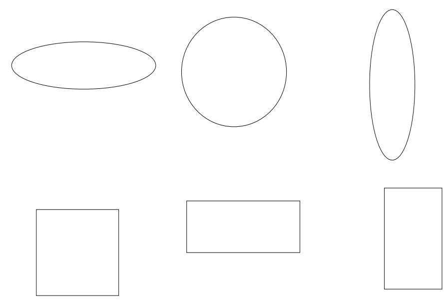

Figure 1: **Ellipses and Rectangles**

Cognition is the act of seeing or perceiving, whereas **recognition** means *as having seen or perceived*. There are three ways of appreciating the activity of **pattern recognition** in a simple manner:

1. **Classification**: Assign a pattern to one of the already known (semantically labelled) classes. For example, consider the two classes of physical objects shown in Figure 1: *ellipses* and *rectangles* where *ellipse* and *rectangle* are class labels. Now the **classification** problem, prominent in pattern recognition, involves:
   (a) Either learn a model or directly use the training data set (collection of labelled patterns) and
   (b) assign a class label to a new pattern (test pattern) or equivalently assign the test pattern to one of the known classes. That is, with respect to objects in Figure 1, given a new object we would like to classify it as either an ellipse or a rectangle.

The Classification problem: We are given a collection of semantically labelled patterns, $\mathcal{X}$, where

$$\mathcal{X} = \{(X_1, C^1), (X_2, C^2), \cdots, (X_n, C^n)\}$$

We need to note the following here:

- The number of classes, K, is fixed and finite. The value of K is known *a priori*. Let the class labels be $C_1, C_2, \cdots, C_K$.
- The set $\mathcal{X}$ is finite and is of size (cardinality) n. Further, X_i represents the i^{th} pattern and C^i is the corresponding semantic class label for $i = 1, \cdots, n$. So, observe that $C^i \in \mathcal{C} = \{C_1, C_2, \cdots, C_K\}$.
- Let X be a test pattern. Then, either we use the training set $\mathcal{X}$ directly or models $M_1, M_2, \cdots, M_K$ learnt from $\mathcal{X}$ to assign a class label, out of $C_1, C_2, \cdots, C_K$, to X. Here, model M_i is learnt from the training patterns drawn from class C_i, for $i = 1, 2, \cdots, K$.

An Example: Let us say that we are given the following collection of *chairs* and *humans* as the training set.

$$\begin{aligned}\mathcal{X} = \{&(X_1, chair), (X_2, chair), (X_3, human), (X_4, human), (X_5, human),\\ &(X_6, chair), (X_7, human), (X_8, chair), (X_9, human), (X_{10}, human)\}\end{aligned}$$

Now the problem is, given a test pattern X, classify X as either *chair* or *human*. In other words, assign one of the two class labels to X.

2. **Clustering**: Assign a pattern to one of the *syntactically labelled classes* or clusters. For example, consider two clusters of patterns, labelled C_1 and C_2. Given a new pattern, assign it to either C_1 or C_2 based on the similarity between the pattern and the collection. Here, the labels are syntactic because we can switch the labels of the two collections without affecting the results. *Clustering* is concerned with grouping of patterns based on similarity. Patterns in a cluster are similar to each other whereas patterns in different clusters are dissimilar.

Clustering Problem: We are given a collection, $\mathcal{X}$, of syntactically labelled patterns, where

$$\mathcal{X} \; = \; \{X_1, \; X_2, \; \cdots, \; X_n\}.$$

Note that the patterns are syntactically labelled using different subscripts. The problem is to partition the set $\mathcal{X}$ into some finite number of blocks or clusters. In other words, we partition $\mathcal{X}$, so that

$$\mathcal{X} \; = \; C_1 \, \cup \, C_2 \, \cup C_3 \, \cdots \, \cup \, C_K$$

where C_i is the i^{th} cluster. Clustering is done so that none of the K clusters is empty and any pair of clusters do not overlap, which means

$$C_i \; \neq \; \emptyset, \; and \; C_i \bigcap C_j \; = \emptyset \; for \; i \neq j \; and \; i, j \; \in \; \{1, 2, \; \cdots, \; K\}.$$

An Example of Clustering: Consider a collection of patterns

$$\mathcal{X} \; = \; \{X_1, \; X_2, \; \cdots, \; X_{10}\}.$$

A possible partition, of $\mathcal{X}$ having two clusters is $C_1 \; = \; \{X_1, X_2, X_4, X_5, X_7, X_8\}$ and $C_2 \; = \; \{X_3, X_6, X_9, X_{10}\}$. Typically, a notion of *similarity* or *matching* is used to partition $\mathcal{X}$. Patterns in C_1 are similar to other patterns in C_1 and patterns in C_2 are similar to other patterns in C_2; a pattern, say X_2, in C_1 is dissimilar to a pattern, say X_9, in C_2. In clustering, it is possible to switch the labels; for example, we have the same partition as above if

$$\begin{aligned} C_1 &= \{X_3, X_6, X_9, X_{10}\} \\ C_2 &= \{X_1, X_2, X_4, X_5, X_7, X_8\} \end{aligned}$$

3. **Semi-Supervised Classification**: Here, we are given a small collection of semantically labelled patterns and a large collection of syntactically labelled patterns. The problem is to assign a new pattern (test pattern) to one of the classes or equivalently assign a semantic label to the test pattern.

Semi-Supervised Classification Problem: We are given a collection, $\mathcal{X}$, given by

$$\mathcal{X} \ = \ \{(X_1, C^1), \ \cdots \ (X_l, C^l), \ X_{l+1}, \ \cdots \ X_{l+u}\}$$

where l patterns are semantically labelled and u patterns are syntactically labelled. The problem is to build models $M_1, \ M_2, \ \cdots \ M_K$ corresponding to classes $C_1, \ C_2, \ \cdots, C_K$ respectively. Now given a new pattern, X, classify it to one of the K classes using the models built.

An Example: Given a set, $\mathcal{X}$, of patterns given by

$$\mathcal{X} \ = \ \{(X_1, human), \ (X_2, chair), \ X_3, \ X_4, \ X_5, \ X_6, \ X_7\}$$

the problem is to assign a class label of *chair* or *human* to a new pattern (test pattern) X.

The popularity of pattern recognition (PR) may be attributed to its application potential; there are several important applications. For example,

- **document recognition:** there are a variety of applications including classification and clustering of
  * email messages and web documents; one requirement is to recognize whether a *mail is spam* or not.
  * fingerprints, face images, and speech signals which form an important variety of documents used in *biometrics*.
  * *health records* which may include x-ray images, ultrasound images, ECG charts and reports on various tests, diagnosis, and prescriptions of medicines.
  * *legal records* including judgments delivered, petitions and appeals made.
- **remote sensed data analysis:** for example, images obtained using satellite or aerial survey are analysed to discriminate healthy crops from deceased crops.
- **bioinformatics:** Here, classification and clustering of DNA and protein sequences is an important activity.

- **semantic computing:** Knowledge in different forms is used in clustering and classification to facilitate natural language understanding, software engineering, and information retrieval.
- There are plenty of other areas like *agriculture, education, and economics* where pattern recognition tools are routinely used.

Abstractions

In machine recognition of patterns, we need to process patterns so that their representations can be stored on the machine. Not only the pattern representations, but also the classes and clusters need to be represented appropriately. In pattern recognition, inputs are abstractions and the outputs also are abstractions.

- As a consequence, we do not need to deal with all the specific details of the individual patterns.
- It is meaningful to summarize the data appropriately or look for an apt abstraction of the data.
- Such an abstraction is friendlier to both the human and the machine.
- For the human it is easy for comprehension and for the machine it reduces the computational burden in the form time and space required for processing.
- Generating an abstraction from examples is a well-known paradigm in machine learning.
- Specifically, learning from examples or supervised learning and learning from observations or clustering are the two important machine learning paradigms that are useful here.
- In artificial intelligence, the machine learning activity is enriched with the help of domain knowledge; abstractions in the form of rule-based systems are popular in this context.
- In addition data mining tools are useful when the set of training patterns is large.

- So, naturally pattern recognition overlaps with machine learning, artificial intelligence and data mining.

Two popular paradigms for pattern recognition are:

- **statistical pattern recognition:** In this case, **vector-spaces** are used to represent patterns and collections of patterns. Vector-space representations are popular in *information retrieval*, *data mining*, and *statistical machine learning*. Abstractions like vectors, graphs, rules or probability distributions are used to represent clusters and classes.

- **syntactic pattern recognition:** In this case, patterns are viewed as sentences in a formal language like mathematical logic. So, it is useful in describing classes and clusters of well-structured patterns. This paradigm is popular as *linguistic* or *structural pattern recognition*.

- Readers interested in some of these applications may refer to popular journals such as Pattern Recognition (www.elsevier.com/locate/pr) and IEEE Transactions on Pattern Analysis and Machine Intelligence (www.computer.org/tpami) for details. Similarly, for specific application areas like bioinformatics refer to Bioinformatics (http://bioinformatics.oxfordjournals.org/) and for semantic computing refer to International Journal of Semantic Computing (www.worldscinet.com/ijsc/). An excellent introduction to syntactic pattern Recognition is providede by *Syntactic Pattern Recognition: An Introduction* by RC Gonzalez and MG Thomason, Addision-Wesley, 1978.

Assignment

Solve the following problems:

1. Consider the data, of four adults indicating their health status, shown in the following table. Devise a simple classifier that can properly classify all the four patterns. How is the fifth adult having a weight of 65 KGs classified using the classifier?

Weight of Adults in KGs	Class label
50	Unhelathy
60	Healthy
70	Healthy
80	Unhealthy

2. Consider the data items bought in a supermarket. The features include cost of the item, size of the item, colour of the object and the class label. The data is shown in the following table. Which feature would you like to use for classification? Why?

item no	cost in Rs.	volume in cm^3	colour	Class label
1	10	6	blue	inexpensive
2	15	6	blue	inexpensive
3	25	6	blue	inexpensive
4	150	1000	red	expensive
5	215	100	red	expensive
6	178	120	red	expensive

Different Paradigms for Pattern Recognition

- There are several paradigms in use to solve the pattern recognition problem.
- The two main paradigms are
  1. Statistical Pattern Recognition
  2. Syntactic Pattern Recognition
- Of the two, the statistical pattern recognition has been more popular and received a major attention in the literature.
- The main reason for this is that most of the practical problems in this area have to deal with noisy data and uncertainty and statistics and probability are good tools to deal with such problems.
- On the other hand, formal language theory provides the background for syntactic pattern recognition. Systems based on such linguistic tools, more often than not, are not ideally suited to deal with noisy environments. However, they are powerful in dealing with well-structured domains. Also, recently there is a growing interest in statistical pattern recognition because of the influence of statistical learning theory.
- This naturally prompts us to orient material in this course towards statistical classification and clustering.

Statistical Pattern Recognition

- In statistical pattern recognition, we use vectors to represent patterns and class labels from a label set.
- The abstractions typically deal with probability density/distributions of points in multi-dimensional spaces, trees and graphs, rules, and vectors themselves.
- Because of the vector space representation, it is meaningful to talk of subspaces/projections and similarity between points in terms of distance measures.

- There are several soft computing tools associated with this notion. Soft computing techniques are tolerant of imprecision, uncertainty and approximation. These tools include neural networks, fuzzy systems and evolutionary computation.
- For example, vectorial representation of points and classes are also employed by
  - neural networks,
  - fuzzy set and rough set based pattern recognition schemes.
- In pattern recognition, we assign labels to patterns. This is achieved using a set of semantically labelled patterns; such a set is called the *training data set*. It is obtained in practice based on inputs from experts.
- In Figure 1, there are patterns of Class 'X' and Class '+'.

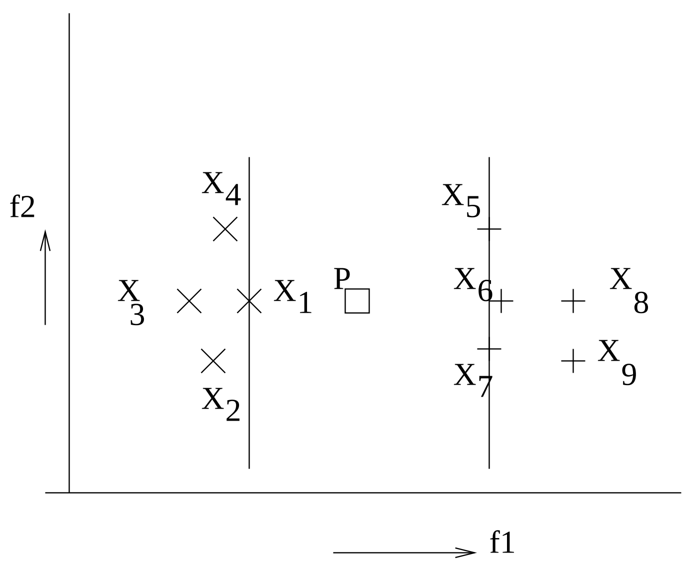

Figure 1: Example set of patterns

- The pattern P is a new sample (test sample) which has to be assigned either to Class 'X' or Class '+'. There are different possibilities; some of them are

  – *The nearest neighbour classifier (NNC)*: Here, P is assigned to the class of its nearest neighbour. Note that pattern X_1 (labelled 'X') is the nearest neighbour of P. So, the test pattern P is assigned the class label 'X'. The nearest neighbour classifier is explained in Module 7.

  – The *K-Nearest neighbour classifier (KNNC)* is based on the class labels of K nearest neighbours of the test pattern P. Note that patterns X_1 (from class 'X'), X_6 (from class '+') and X_7 (from class '+') are the first three (K=3) neighbours.A majority (2 out of 3) of the neighbours are from class '+'. So, P is assigned the class label '+'. We discuss the KNNC in module 7.

  – *Decision stump classifier:* In this case, each of the two features is considered for splitting; the one which provides the best separation between the two classes is chosen. The test pattern is classified based on this split. So, in the example, the test pattern P is classified based on whether its first feature (x-coordinate) value is less than A or not. If it is less than A, then the class is 'X', else it is '+'. In Figure 1, P is assigned to class 'X'. A generalization of the decision stump called the *decision tree classifier* is studied in module 12.

  – *Separating line as decision boundary:* In Figure 1, the two classes may be characterized in terms of the boundary patterns falling on the support lines. In the example, pattern X_1 (class 'X') falls on one line (say line1) and patterns X_5 and X_7 (of class '+') fall on a parallel line (line2). So, any pattern closer to line 1 is assigned the class label 'X' and similarly patterns closer to line2 are assigned class label '+'. We discuss classifiers based on such linear discriminants in module 12. Neural networks and support vector machines (SVMs) are members of this category. We discuss them in module 13.

  – It is possible to use a *combinations of classifiers* to classify a test pattern. For example, P could be classified using weighted nearest neighbours. Suppose such a weighted classifier assigns a weight of 0.4 to the first neighbour (pattern X_1, labelled 'X'), a weight of 0.35 to the second neighbour (pattern X_6 from class '+') and a weight of 0.25 to the third neighbour (pattern X_7 from class '+'). We first add the weights of the neighbours of P coming from the same class. So, the sum of the weights for class 'X', W_X is 0.4 as only the first neighbour is from 'X'. The sum of the weights for class '+', W_+ is 0.6 (0.35 $`+'$ 0.25) corresponding the remaining two neighbours (8 and 6) from class '+'. So, P is assigned class label '+'. We discuss combinations of classifiers in module 16.

- In a system that is built to classify humans into tall, medium and short, the abstractions, learnt from examples, facilitate assigning one of these class labels (tall, medium or short) to a newly encountered human. Here, the class labels are semantic; they convey some meaning.
- In the case of clustering, we can group a collection of unlabelled patterns also; in such a case, the labels assigned to each group of patterns is syntactic, simply the cluster identity.
- Several times, it is possible that there is a large training data which can be directly used for classification. In such a context, clustering can be used to generate abstractions of the data and use these abstractions for classification. For example, sets of patterns corresponding to each of the classes can be clustered to form subclasses. Each such subclass (cluster) can berepresented by a single prototypical pattern; these representative patterns can be used to build the classifier instead of the entire data set. In Modules 14 and 15, a discussion on some of the popular clustering algorithms is presented.

Importance of Representation

- It is possible to directly use a classification rule without generating any abstraction, for example by using the NNC.
- In such a case, the notion of proximity/similarity (or distance) is used to classify patterns.

- Such a similarity function is computed based on the representation of patterns; the representation scheme plays a crucial role in classification.
- A pattern is represented as a vector of feature values.
- The features which are used to represent patterns are important. We illustrate it with the help of the following example.

Example

Consider the following data where humans are to be categorized into tall and short. The classes are represented using the feature Weight. If a newly

Weight of human (in Kilograms)	Class label
40	tall
50	short
60	tall
70	short

encountered person weighs 46 KGs, then he/she may be assigned the class label short because 46 is closer to 50. However, such an assignment does not appeal to us because we know that weight and the class labels *tall* and *short* do not correlate well; a feature such as Height is more appropriate. Module 2 deals with representation of patterns and classes.

Overview of the course

- Modules 3-6 deal with representation of patterns and classes. Also, proximity between patterns is discussed in these modules.
- Various classifiers are discussed in modules 7 to 13 and module 16.
  - The most popular and simple classifier is based on the NNC. In such a classification scheme, we do not have any training phase. A detailed discussion on nearest neighbor classification is presented in Module 7, 8, and 9.

  - It is important to look for theoretical aspects of the limits of classifiers under uncertainty. Bayes classifier characterizes optimality in terms of minimum error-rate classification. It is discussed in Module 10.
  - A decision tree is a transparent data structure to deal with classification of patterns employing both numerical and categorical features. We discuss decision tree classifiers in Module 11.
  - Using linear decision boundaries in high-dimensional spaces has gained a lot of prominence in the recent past. Support vector machines (SVMs) are built based on this notion. In Module 12 and 13, the role of SVMs in classification is explored.
  - It is meaningful to use more than one classifier to arrive at the class label of a new pattern. Such combination of classifiers forms the basis for Module 16.

- In Modules 14 a discussion on some of the popular clustering algorithms is presented.
- There are several challenges faced while clustering large datasets. In module 15 some of these challenges are outlined and algorithms for clustering large datasets are presented.
- Finally we consider an application of document classification and retrieval in module 17.

Assignment

1. Consider a collection of data items bought in a supermarket. The features include cost of the item, size of the item and the class label. The data is shown in the following table. Consider a new item with cost = 34 and volume = 8. How do you classify this item using the NNC? How about KNNC with K = 3?
2. Consider the problem of classifying objects into *triangles* and *rectangles*. Which paradigm do you use? Provide an appropriate representation.
3. Consider a variant of the previous problem where the classes are *small circle* and *big circle*. How do you classify such objects?

item no	cost in Rs.	volume in cm^3	Class label
1	10	6	inexpensive
2	15	6	inexpensive
3	25	6	inexpensive
4	50	10	expensive
5	45	10	expensive
6	47	12	expensive

Further Reading

[1] is an introductory book on Pattern Recognition with several worked out examples. [2] is an excellent book on Pattern Classification. [5] is a book on data mining. [3] is an book on artificial intelligence which discusses learning and pattern recognition techniques as a part of artificial intelligence. Neural network as used for Pattern Classification is found in [4].

References

[1] V. Susheela Devi, M. Narasimha Murty. *Pattern Recognition: An Introduction* Universities Press, Hyderabad, 2011.

[2] R.O. Duda, P.E. Hart, D.G. Stork. Pattern Classification John Wiley and Sons, 2000.

[3] S. Russell and P. Norvig *Artificial intelligence A Modern approach* Pearson India, 2003.

[4] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, New Delhi, 2003.

[5] P. N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining* Pearson India, 2007.

What is a pattern?

- A pattern represents a physical object or an abstract notion. For example, the pattern may represent physical objects like balls, animals or furniture. Abstract notions could be like whether a person will play tennis or not(depending on features like weather etc.).
- It gives the description of the object or the notion.
- The description is given in the form of attributes of the object.
- These are also called the features of the object.

What are classes?

- The patterns belong to two or more classes.
- The task of pattern recognition pertains to finding the class to which a pattern belongs.
- The attributes or features used to represent the patterns should be discriminatory attributes. This means that they help in classifying the patterns.
- The task of finding the discriminatory features is called feature extraction/selection.

What is classification?

- Given a pattern, the task of identifying the class to which the pattern belongs is called classification.
- Generally, a set of patterns is given where the class label of each pattern is known. This is known as the training data.
- The information in the training data should be used to identify the class of the test pattern.

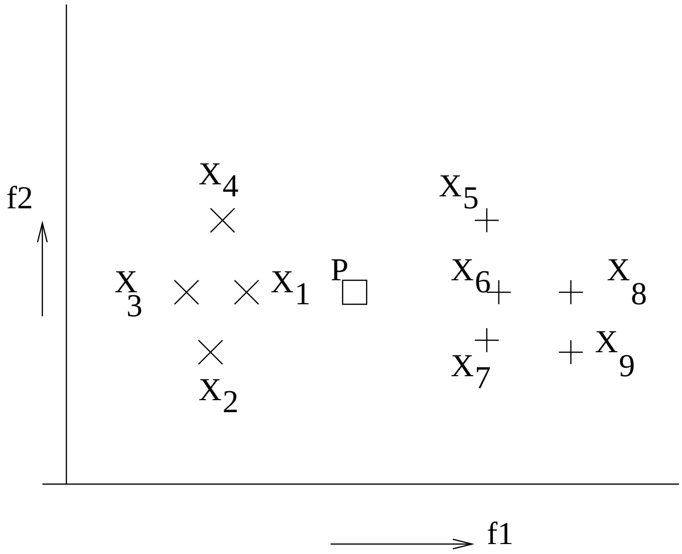

Figure 1: Dataset of two classes

- This type of classification where a training set is used is called supervised learning. In supervised learning, we can learn about the values of the features for each class from the training set and using this information, a given pattern is classified.

  Consider the patterns of two classes given in Figure 1. This is the training data.

  Using the training data, we can classify the pattern P. The information of the two classes available in the training data can be used to carry out this classification. There are a number of classifiers which carry out supervised classification like nearest neighbour and related algorithms, Bayes classifier, decision trees, SVM, neural networks, etc which are discussed in later modules.

Representation of patterns

- Patterns can be represented in a number of ways.

- All the ways pertains to giving the values of the features used for that particular pattern.

- For supervised learning, where a training set is given, each pattern in the training set will also have the class of the pattern given.

Representing patterns as vectors

- The most popular method of representing patterns is as vectors.
- Here, the training dataset may be represented as a matrix of size (nxd), where each row corresponds to a pattern and each column represents a feature.
- Each attribute/feature/variable is associated with a domain. A domain is a set of numbers, each number pertains to a value of an attribute for that particular pattern.
- The class label is a dependent attribute which depends on the 'd' independent attributes.

Example

The dataset could be as follows :

	f_1	f_2	f_3	f_4	f_5	f_6	Class label
Pattern 1:	1	4	3	6	4	7	1
Pattern 2:	4	7	5	7	4	2	2
Pattern 3:	6	9	7	5	3	1	3
Pattern 4:	7	4	6	2	8	6	1
Pattern 5:	4	7	5	8	2	6	2
Pattern 6:	5	3	7	9	5	3	3
Pattern 7:	8	1	9	4	2	8	3

In this case, n=7 and d=6. As can be seen,each pattern has six attributes(or features). Each attribute in this case is a number between 1 and 9. The last number in each line gives the class of the pattern. In this case, the class of the patterns is either 1, 2 or 3.

- If the patterns are two- or three-dimensional, they can be plotted.
- Consider the dataset

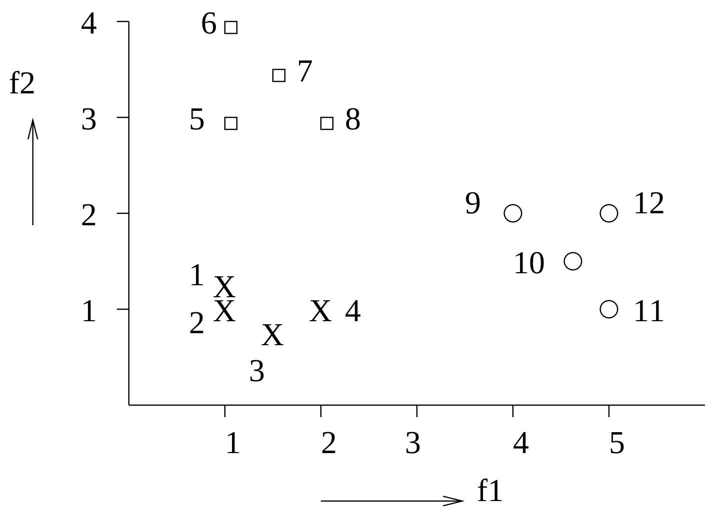

Figure 2: Dataset of three classes

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)

Each triplet consists of feature 1, feature 2 and the class label. This is shown in Figure 2.

Representing patterns as strings

- Here each pattern is a string of characters from an alphabet.
- This is generally used to represent gene expressions.
- For example, DNA can be represented as

  GTGCATCTGACTCCT...

  RNA is expressed as

GUGCAUCUGACUCCU....

This can be translated into protein which would be of the form

VHLTPEEK ....

- Each string of characters represents a pattern. Operations like pattern matching or finding the similarity between strings are carried out with these patterns.
- More details on proteins and genes can be got from [1].

Representing patterns by using logical operators

- Here each pattern is represented by a sentence(well formed formula) in a logic.
- An example would be
  if (beak(x) = red) and (colour(x) = green) then parrot(x)
  This is a rule where the antecedent is a conjunction of primitives and the consequent is the class label.
- Another example would be
  if (has-trunk(x)) and (colour(x) = black) and (size(x) = large) then elephant(x)

Representing patterns using fuzzy and rough sets

- The features in a fuzzy pattern may consist of linguistic values, fuzzy numbers and intervals.
- For example, linguistic values can be like tall, medium, short for height which is very subjective and can be modelled by fuzzy membership values.

- A feature in the pattern maybe represented by an interval instead of a single number. This would give a range in which that feature falls. An example of this would be the pattern

  $(3, small, 6.5, [1, 10])$

  The above example gives a pattern with 4 features. The 4th feature is in the form of an interval. In this case the feature falls within the range 1 to 10. This is also used when there are missing values. When a particular feature of a pattern is missing, looking at other patterns, we can find a range of values which this feature can take. This can be represented as an interval.

  The example pattern given above has the second feature as a linguistic value. The first feature is an integer and the third feature is a real value.

- Rough sets are used to represent classes. So, a class description will consist of an upper approximate set and a lower approximate set. An element y belongs to the lower approximation if the equivalence class to which y belongs is included in the set. On the other hand y belongs to the upper approximation of the set if its equivalence class has a non-empty intersection with the set. The lower approximation consists of objects which are members of the set with full certainty. The upper approximation consists of objects which may possibly belong to the set.

- For example, consider Figure 3. This represents an object whose location can be found by the grid shown. The object shown completely covers (A3,B2), (A3,B3), (A4,B2) and (A4,B3). The object falls partially in (A2,B1),(A2,B2),(A2,B3), (A2,B4),(A3,B1),(A3,B4),(A4,B1),(A4,B4), (A5,B2), and (A5,B3). The pattern can be represented as a rough set where the first four values of the grid gives the lower approximation and the rest of the values of the grid listed above form the upper approximation.

- Not just the features, each pattern can have grades of membership to every class instead of belonging to one class. In other words, each

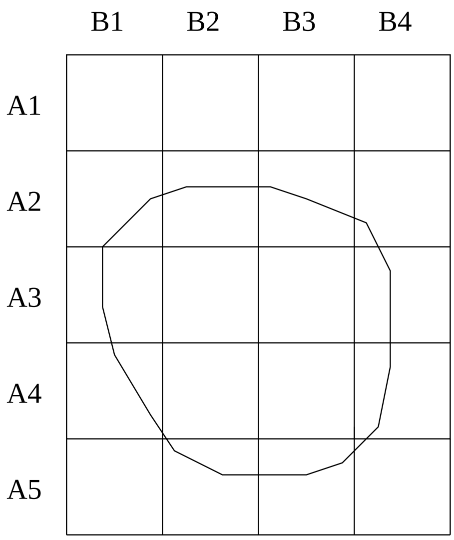

Figure 3: Representation of an object

pattern has a fuzzy label which consists of c values in [0,1] where each component gives the grade of membership of the pattern to one class. Here c gives the number of classes. For example, consider a collection of documents. It is possible that each of the documents may be associated with more than one category. A paragraph in a document, for instance, may be associated with *sport* and another with *politics*.

- The classes can also be fuzzy. One example of this would be to have linguistic values for classes. The classes for a set of patterns can be *small* and *big*. These classes are fuzzy in nature as the perception of small and big is different for different people.

References

[1] Andreas D. Baxevanis(Ed), B.F. Francis Ouelette(Ed) Bioinformatics : A Practical Guide to the Analysis of Genes and Proteins John Wiley and Sons Incorporated, 3rd Edition, October 2004.