MODULE 2
Paradigms for Pattern Recognition

LESSON 2
Statistical Pattern Recognition

Keywords: Statistical, Syntactic, Representation, Vector Space, Classification
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 be represented 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 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.

<table>
<thead>
<tr>
<th>Weight of human (in Kilograms)</th>
<th>Class label</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>tall</td>
</tr>
<tr>
<td>50</td>
<td>short</td>
</tr>
<tr>
<td>60</td>
<td>tall</td>
</tr>
<tr>
<td>70</td>
<td>short</td>
</tr>
</tbody>
</table>

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?
<table>
<thead>
<tr>
<th>item no</th>
<th>cost in Rs.</th>
<th>volume in $cm^3$</th>
<th>Class label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>6</td>
<td>inexpensive</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>6</td>
<td>inexpensive</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>6</td>
<td>inexpensive</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>10</td>
<td>expensive</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>10</td>
<td>expensive</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>12</td>
<td>expensive</td>
</tr>
</tbody>
</table>

**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**


