Preface

Pattern recognition is a very general concept to insure that the process of recognition and decision in complex situations becomes automated. It is an area of work that has applications in many disciplines. It is expected, that the reader has to deal with the digital signal analysis and with different methods at the feature extraction procedures. At the classification and cluster analysis procedures he has to have a good look at the concepts of statistics and matrix calculation. The formal representation uses the language of mathematics, but all intellectual steps are justified by verbal argumentation. Many examples help the reader to understand the written pages and to try them out for themselves. A scanner e. g. can digitize characters and recognize them by different methods. MATLAB offers possibilities recording data using a microphone. With these data a speech recognition system can be designed.

1.0 Introduction

A pattern recognition system describes tasks, at which measured objects shall be assigned to certain classes. The measurements decide which class the object belongs to. The measurements or their derived sizes are described generally as feature vectors.

1.1 Where can you use pattern recognition?

- recognition of characters

  measuring: two-dimensional signal f(x, y) grey -values
  decision: character, e.g. A, B
  applications: reading letter addresses,
  reading bank documents
  devices to read to blind people

- speech recognition

  measuring: one-dimensional signal f(t)
  output voltage of a microphone
  derived energy in certain frequency domains,
  pitch frequency
  decision: language sound or word
  applications: package sorting, telephone advice services,
  final inspection of cars,
  general examining tasks
  input 10-100 words, at present still speaker dependent procedures

- speaker recognition
  - measuring: see speech recognition
  - decision: speaker identification and speaker verification. At a speaker identification you have n speakers with their voices known. The task is to find out to whom the unknown voice belongs. The speaker verification is used for an unknown speaker to verify his identity inspecting his voice. The task is to find out if the person belongs to the stored voice.
  - Applications: access control to a computing room, computer data base, nuclear power stations, goods order, transactions in a bank, police identification
- medical diagnosis
  - measuring: e.g. blood pressure, blood sugar, pulse ....
  - decision: diagnosis, e.g. hypertension, diabetes
- machine tools
  - measuring: acceleration signals, ultrasonic signals, directional signals, strength signals, torque
  - decision: leakage and breakage or abrasion test for tools
  - applications: process monitoring for drilling, milling, sawing, turning and so on

The classifier only has to derive a decision on the class membership analysing the measuring. The essential difficulty of pattern recognition consists of the fact, that patterns of the same class are not characterized by one and the same feature vector. There is a complicated connection between the measuring and the true class membership.

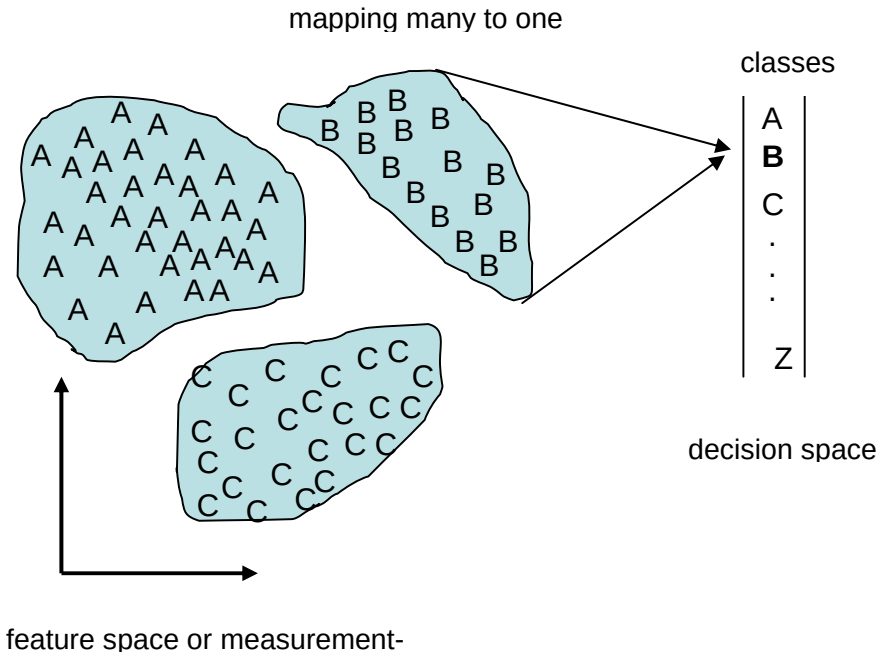

Figure 1.01: Transformation or mapping from the feature space into the decision space

You can get access to the statistical laws which describe the true class membership from the set of all patterns by a simultaneous statement of the class membership (learning set). Apart from the learning set, which the mapping from measurement space to decision space is learned from, you need a test set to check the quality of the classifier.

1.2 Feature extraction and classification.

The purpose of the feature extraction is to reduce the features without losing the information important for the class membership. The procedure of checking the system performance based on the learning set is often called ***reclassification error***, and the procedure of checking it with an independent test set is called ***generalisation error***. Relevant for application is the generalisation error.

Example: speech

A spoken word of 1 sec duration consists of 8000 values (time domain) at a sampling frequency of 8 kHz. If you calculate the periodogram you get only 40 values (frequency domain). The less features you have, the simpler is the computation of the classifier data, but the generalisation error is still the most important.

There are innumerable procedures for feature extraction:

- heuristic method: features are defined by intuition and experience.
- analytical method: With the use of a criterion, e.g. minimal reconstruction error you can try to derive an optimal feature set systematically.
- choice of features: The feature vector is divided because of too many features. With the help of the least squares method classifier or other methods [Narenda77] a sequence list of the feature is determined and step by step 5, 10 ⋯features are used to calculate the error rate.

Conditions which should be fulfilled in practice:

Extensive learning set:

For the extraction of the features and calculating the classifier data many samples are necessary. To calculate a good classifier you need a representative learning set. Text books often present data describing how many feature vectors you need to calculate a classifier.
In practice it is shown that these data are often incorrect.
If the learning set is classified, you can call this the reclassification error rate. An additional use of an extensive test set is advised.

Extensive test set:

Independently of the learning set error rate the test set error rate gives information whether separation effective features were found or whether the test classification error rate corresponds to the requirements.
Figure 1.02 represents the connection of the error rates. In practice the error rate ε_0 rarely lowers towards zero. But the error rate zero shows an over adaptation with a finite learning set. The consequence is an increased test classification error rate.

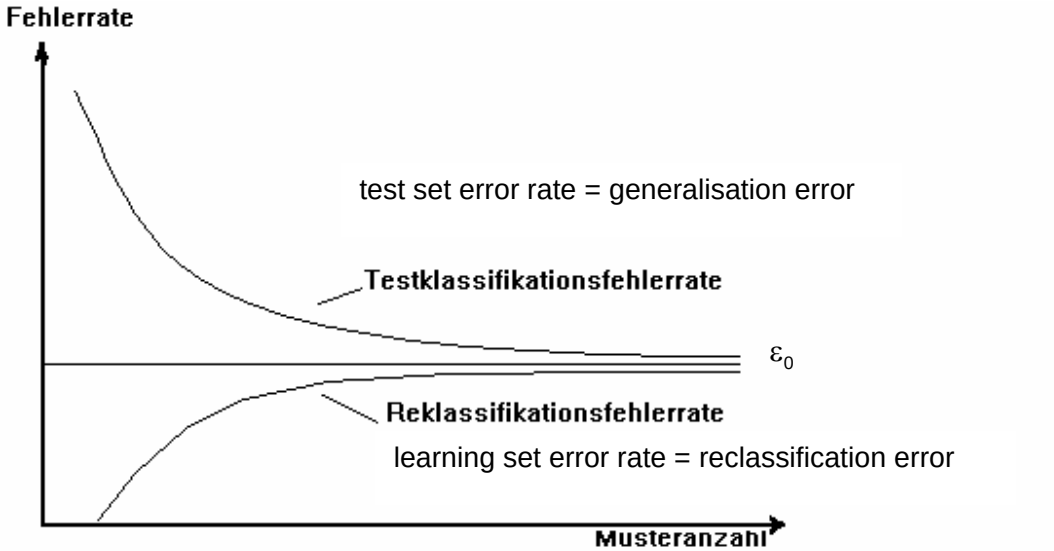

Figure 1.02: Connection between the learning set error rate and the test set error rate, depending on the number of patterns.

Pearl of wisdom:
A reclassification error rate that has not enough feature vectors, is completely insignificant. The generalisation error proves itself as a good quality measure for the separation of the features.
Figure 1.03 shows a general classification system. The features have to be selected so that the decision always falls for the "right" class or to put it into different words the classification error shall be minimized. There is no general algorithm for the minimization of the classification error.

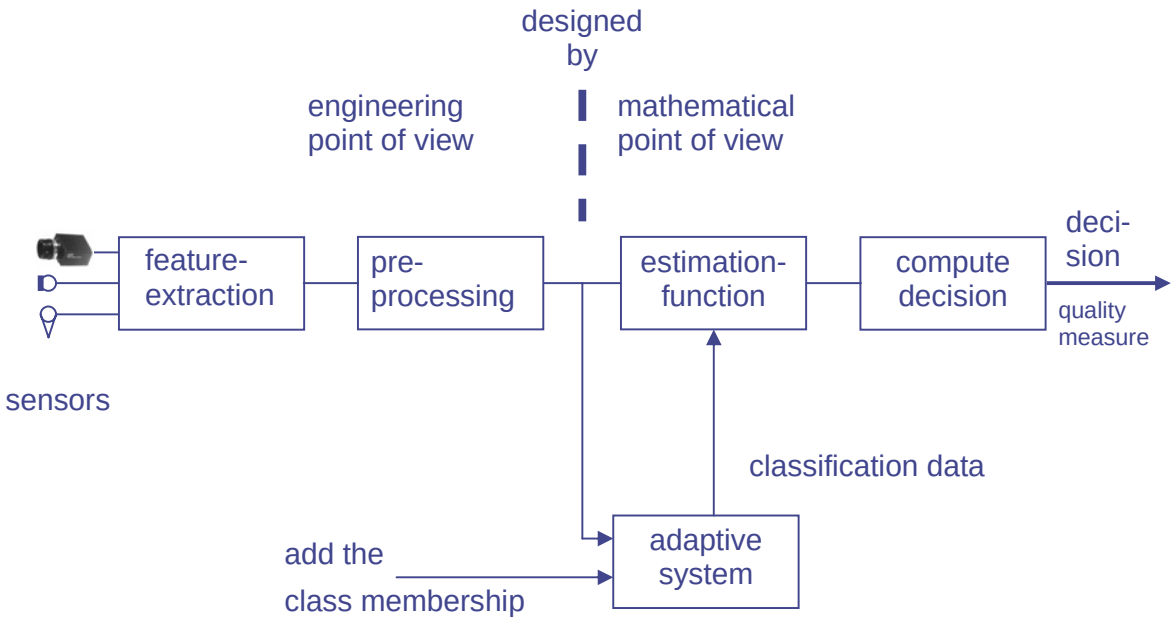

Figure 1.03: Classification system

Feature extraction is carried out under engineering points of view and determines the "success" of the classifier. We have a lot of mathematical optimization methods, which hardly leave room to develop a new classification system. On the other side the feature extraction is an unlimited area for the developer. The vectors of the learning set have to be combined with the true class membership and thereafter the classifier data can be calculated.

As soon as the classifier data are calculated and taken into the estimation function, you can leave out the lower part of figure 1.03. The measurements are taken by the sensors and pre-processed. The estimation function then calculates the probability, to which class the measurements belong. At the same time a quality measure is computed, which allows a statement about the reliability of the classification. You can get some MATLAB programs for different classification procedures, look: www.hit-karlsruhe.de/becker.

1.3 Summary

The ability of people to recognize a person by their voice or to know the class of an incomplete character motivates professionals and others to develop systems doing exactly the same. Commercial applications like the OCR character recognition are well executed. But it is necessary to show the limits that exist if one likes to solve such questions by the personal computer only.
This chapter introduced you to the complex idea of the pattern recognition, the feature extraction, the class membership and the task of a classifier. At the same time the terms “learning set” and “test set” have been explained. The whole script is dealing with the feature extraction procedures, the statistical classification procedures and the cluster analysis. At the end we are going to deal with image processing as well.