CSIS8502

1. Introduction to Pattern Classification
What is Pattern Classification?

1. Input: Pattern description
   - spacial – characters, fingerprints, etc.
   - temporal – speech, ECG etc.

2. Output: some kinds of classification, e.g.
   - which character is written?
   - whose fingerprint it is?
   - what is spoken?
   - is the ECG normal?
How Pattern Classification is done?

Two approaches:

1. Decision-Theoretic Approach (Statistical Approach)
   - Represent the pattern as a vector in a vector space (*feature space*)
   - Use a decision algorithm (mainly statistical) to decide which class the pattern is in.

2. Structural Approach (Syntactic Approach)
   - Represent the pattern by its structure, e.g. a string of symbols, a graph connecting the primary elements, etc.
   - Use parsing (grammatical) or graph matching to perform classification.
How Pattern Classification is done?

- **Problem-dependent part:**
  1. How to convert the pattern to a point (vector) in the feature space – *feature extraction*.
  2. How to represent the pattern by your desired data structure (string or graph)

- **Problem-independent part:**
  1. decision algorithm, e.g. Classifier etc.
  2. parsing algorithm, graph matching algorithms, etc.
Recognizing *salmon* vs *sea bass*:

- First collect a lot of salmon and sea bass
- Identify the *features* that can differentiate between the two species, e.g. length and lightness.
- Find the distribution (or the histogram) of the features.
- Each fish can then be represented as a 2-dim feature vector, and can be represented as a point in the 2-dim feature space.
- Recognition can then be performed by finding a decision boundary in the feature space.
FIGURE 1.1. The objects to be classified are first sensed by a transducer (camera), whose signals are preprocessed. Next the features are extracted and finally the classification is emitted, here either “salmon” or “sea bass.” Although the information flow is often chosen to be from the source to the classifier, some systems employ information flow in which earlier levels of processing can be altered based on the tentative or preliminary response in later levels (gray arrows). Yet others combine two or more stages into a unified step, such as simultaneous segmentation and feature extraction. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
Figure 1.2. Histograms for the length feature for the two categories. No single threshold value of the length will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value marked \( l^* \) will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. Copyright © 2001 by John Wiley & Sons, Inc.
FIGURE 1.3. Histograms for the lightness feature for the two categories. No single threshold value $x^*$ (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value $x^*$ marked will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
**Figure 1.4.** The two features of lightness and width for sea bass and salmon. The dark line could serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
The decision boundary will classify most of the samples correctly, but may not be 100%.

When a new fish is seen, Generalization is made:
- features are extracted
- the fish is represented as a point in the feature space
- classify according to which area the point lies in.

Theoretically speaking, we can find a complex decision surface that can correctly classify all samples, but this may not be a good generalization, i.e. may not perform well when classifying unseen samples.
FIGURE 1.5. Overly complex models for the fish will lead to decision boundaries that are complicated. While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns. The novel test point marked \( ? \) is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be classified as a sea bass. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
**FIGURE 1.6.** The decision boundary shown might represent the optimal trade-off between performance on the training set and simplicity of classifier, thereby giving highest accuracy on new patterns. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
Assumptions

- We have to make assumptions on:
  - distribution of the features, e.g. assume normal distribution
  - the shape of the decision surface, e.g. assume linear, or quadratic decision surface.
Pattern Recognition Systems

- **Sensing**: use sensor to capture the required pattern image.

- **Segmentation**: segment the objects from the image, e.g. segment the salmon or sea bass from the image captured.

- **Feature extraction**: measure the required features from the object, e.g. length, and lightness. You may need to consider features that possess some special characteristics, e.g. rotational and translational invariant.

- **Classification**: find the decision boundary or decision algorithm.

- **Postprocessing**: adjust the recognition result by looking at other factors, such as the context in which the object is in, or the risk associating with incorrect decision, e.g. in medical application, where risk is very high.
**FIGURE 1.7.** Many pattern recognition systems can be partitioned into components such as the ones shown here. A sensor converts images or sounds or other physical inputs into signal data. The segmentor isolates sensed objects from the background or from other objects. A feature extractor measures object properties that are useful for classification. The classifier uses these features to assign the sensed object to a category. Finally, a post processor can take account of other considerations, such as the effects of context and the costs of errors, to decide on the appropriate action. Although this description stresses a one-way or “bottom-up” flow of data, some systems employ feedback from higher levels back down to lower levels (gray arrows). From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.
The Design Cycle

- **Data collection**: collect a lot of samples of salmons and sea bass for analysis.
- **Feature Choice**: choose the features that can differentiate the two species. (problem dependent)
- **Model Choice**: decide what assumption (model) we want to make, e.g. assuming Normal distribution, or linear decision boundary.
- **Training**: From the data collected, find the best model which suits the training data.
- **Evaluation**: evaluate the performance of the system using unseen data.
FIGURE 1.8. The design of a pattern recognition system involves a design cycle similar to the one shown here. Data must be collected, both to train and to test the system. The characteristics of the data impact both the choice of appropriate discriminating features and the choice of models for the different categories. The training process uses some or all of the data to determine the system parameters. The results of evaluation may call for repetition of various steps in this process in order to obtain satisfactory results. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.