Introduction to Pattern Recognition
Part I

Selim Aksoy

Department of Computer Engineering
Bilkent University
saksoy@cs.bilkent.edu.tr

CS 484, Spring 2011
Humans have developed highly sophisticated skills for sensing their environment and taking actions according to what they observe, e.g.,

- recognizing a face,
- understanding spoken words,
- reading handwriting,
- distinguishing fresh food from its smell.

We would like to give similar capabilities to machines.
What is Pattern Recognition?

- A *pattern* is an entity, vaguely defined, that could be given a name, e.g.,
  - fingerprint image,
  - handwritten word,
  - human face,
  - speech signal,
  - DNA sequence,
  - ...

- *Pattern recognition* is the study of how machines can
  - observe the environment,
  - learn to distinguish patterns of interest,
  - make sound and reasonable decisions about the categories of the patterns.
Table 1: Example pattern recognition applications.

<table>
<thead>
<tr>
<th>Problem Domain</th>
<th>Application</th>
<th>Input Pattern</th>
<th>Pattern Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document image analysis</td>
<td>Optical character recognition</td>
<td>Document image</td>
<td>Characters, words</td>
</tr>
<tr>
<td>Document classification</td>
<td>Internet search</td>
<td>Text document</td>
<td>Semantic categories</td>
</tr>
<tr>
<td>Document classification</td>
<td>Junk mail filtering</td>
<td>Email</td>
<td>Junk/non-junk</td>
</tr>
<tr>
<td>Multimedia database retrieval</td>
<td>Internet search</td>
<td>Video clip</td>
<td>Video genres</td>
</tr>
<tr>
<td>Speech recognition</td>
<td>Telephone directory assistance</td>
<td>Speech waveform</td>
<td>Spoken words</td>
</tr>
<tr>
<td>Natural language processing</td>
<td>Information extraction</td>
<td>Sentences</td>
<td>Parts of speech</td>
</tr>
<tr>
<td>Biometric recognition</td>
<td>Personal identification</td>
<td>Face, iris, fingerprint</td>
<td>Authorized users for access control</td>
</tr>
<tr>
<td>Medical</td>
<td>Computer aided diagnosis</td>
<td>Microscopic image</td>
<td>Cancerous/healthy cell</td>
</tr>
<tr>
<td>Military</td>
<td>Automatic target recognition</td>
<td>Optical or infrared image</td>
<td>Target type</td>
</tr>
<tr>
<td>Industrial automation</td>
<td>Printed circuit board inspection</td>
<td>Intensity or range image</td>
<td>Defective/non-defective product</td>
</tr>
<tr>
<td>Industrial automation</td>
<td>Fruit sorting</td>
<td>Images taken on a conveyor belt</td>
<td>Grade of quality</td>
</tr>
<tr>
<td>Remote sensing</td>
<td>Forecasting crop yield</td>
<td>Multispectral image</td>
<td>Land use categories</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Sequence analysis</td>
<td>DNA sequence</td>
<td>Known types of genes</td>
</tr>
<tr>
<td>Data mining</td>
<td>Searching for meaningful patterns</td>
<td>Points in multidimensional space</td>
<td>Compact and well-separated clusters</td>
</tr>
</tbody>
</table>
An Example

- Problem: Sorting incoming fish on a conveyor belt according to species.
- Assume that we have only two kinds of fish:
  - sea bass,
  - salmon.

Figure 1: Picture taken from a camera.
An Example: Decision Process

- What kind of information can distinguish one species from the other?
  - length, width, weight, number and shape of fins, tail shape, etc.

- What can cause problems during sensing?
  - lighting conditions, position of fish on the conveyor belt, camera noise, etc.

- What are the steps in the process?
  - capture image → isolate fish → take measurements → make decision
An Example: Selecting Features

- Assume a fisherman told us that a sea bass is generally longer than a salmon.
- We can use length as a feature and decide between sea bass and salmon according to a threshold on length.
- How can we choose this threshold?
An Example: Selecting Features

Figure 2: *Histograms* of the length feature for two types of fish in *training samples*. How can we choose the threshold $l^*$ to make a reliable decision?
An Example: Selecting Features

► Even though sea bass is longer than salmon on the average, there are many examples of fish where this observation does not hold.

► Try another feature: average lightness of the fish scales.
An Example: Selecting Features

Figure 3: Histograms of the lightness feature for two types of fish in training samples. It looks easier to choose the threshold $x^*$ but we still cannot make a perfect decision.
We should also consider \textit{costs of different errors} we make in our decisions.

For example, if the fish packing company knows that:

- Customers who buy salmon will object vigorously if they see sea bass in their cans.
- Customers who buy sea bass will not be unhappy if they occasionally see some expensive salmon in their cans.

How does this knowledge affect our decision?
An Example: Multiple Features

- Assume we also observed that sea bass are typically wider than salmon.

- We can use two features in our decision:
  - lightness: $x_1$
  - width: $x_2$

- Each fish image is now represented as a point (feature vector)

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

in a two-dimensional feature space.
An Example: Multiple Features

Figure 4: Scatter plot of lightness and width features for training samples. We can draw a decision boundary to divide the feature space into two regions. Does it look better than using only lightness?
An Example: Multiple Features

- Does adding more features always improve the results?
  - Avoid unreliable features.
  - Be careful about correlations with existing features.
  - Be careful about measurement costs.
  - Be careful about noise in the measurements.
- Is there some *curse* for working in very high dimensions?
An Example: Decision Boundaries

- Can we do better with another decision rule?
- More complex models result in more complex boundaries.

Figure 5: We may distinguish training samples perfectly but how can we predict how well we can generalize to unknown samples?
An Example: Decision Boundaries

- How can we manage the tradeoff between complexity of decision rules and their performance to unknown samples?

Figure 6: Different criteria lead to different decision boundaries.
Figure 7: Object/process diagram of a pattern recognition system.
Pattern Recognition Systems

- Data acquisition and sensing:
  - Measurements of physical variables.
  - Important issues: bandwidth, resolution, sensitivity, distortion, SNR, latency, etc.

- Pre-processing:
  - Removal of noise in data.
  - Isolation of patterns of interest from the background.

- Feature extraction:
  - Finding a new representation in terms of features.
Pattern Recognition Systems

- Model learning and estimation:
  - Learning a mapping between features and pattern groups and categories.

- Classification:
  - Using features and learned models to assign a pattern to a category.

- Post-processing:
  - Evaluation of confidence in decisions.
  - Exploitation of context to improve performance.
  - Combination of experts.
The Design Cycle

Collect data → Select features → Select model → Train classifier → Evaluate classifier

Figure 8: The design cycle.

Data collection:
- Collecting training and testing data.
- How can we know when we have adequately large and representative set of samples?
The Design Cycle

- Feature selection:
  - Domain dependence and prior information.
  - Computational cost and feasibility.
  - Discriminative features.
    - Similar values for similar patterns.
    - Different values for different patterns.
  - Invariant features with respect to translation, rotation and scale.
  - Robust features with respect to occlusion, distortion, deformation, and variations in environment.
The Design Cycle

Model selection:

- Domain dependence and prior information.
- Definition of design criteria.
- Parametric vs. non-parametric models.
- Handling of missing features.
- Computational complexity.
- Types of models: templates, decision-theoretic or statistical, syntactic or structural, neural, and hybrid.
- How can we know how close we are to the true model underlying the patterns?
Training:

- How can we learn the rule from data?
- Supervised learning: a teacher provides a category label or cost for each pattern in the training set.
- Unsupervised learning: the system forms clusters or natural groupings of the input patterns.
- Reinforcement learning: no desired category is given but the teacher provides feedback to the system such as the decision is right or wrong.
The Design Cycle

► Evaluation:
  ▶ How can we estimate the performance with training samples?
  ▶ How can we predict the performance with future data?
  ▶ Problems of overfitting and generalization.
Summary

- Pattern recognition techniques find applications in many areas: machine learning, statistics, mathematics, computer science, biology, etc.
- There are many sub-problems in the design process.
- Many of these problems can indeed be solved.
- More complex learning, searching and optimization algorithms are developed with advances in computer technology.
- There remain many fascinating unsolved problems.