# ICS663: Pattern Recognition

Department of Information and Computer Sciences
University of Hawai`I at Manoa

Kyungim Baek



## Lecture 1

- Overview
  - Applications of Pattern Recognition
  - Classification Example
  - Components of Pattern Recognition Systems
  - Learning and Adaptation in Pattern Recognition



## A Recognition Problem

- Male… or… Female?

Slide courtesy of S. Iliescu



## Pattern Recognition

- The assignment of a physical object or event to one of several pre-specified categories (Duda & Hart)
- Concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories (C. Bishop)
- The bulk of human intelligence is based on pattern recognition: the quintessential example of self-organization

## Related Fields of Study

- Signal processing
- Machine learning
- Artificial neural networks
- Robotics and vision
- Cognitive science
- Computational neuroscience
- Mathematical statistics
- Nonlinear optimization
- Exploratory data analysis
- Fuzzy and genetic algorithms
- Detection and estimation theory
- Formal languages
- etc…



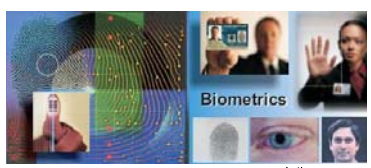
## Growing Applications of PR

- Speech recognition
- Automated target recognition
- Optical character recognition
- Seismic analysis
- Biometrics
- Medical diagnosis
- Data mining
- Gene sequence analysis
- Aerial reconnaissance
- Human computer interaction
- etc…



## Biometrics

- Human identification based on biological patterns
- Static pattern
  - Fingerprint, iris, face, palm print, hand geometry, vein of the back of the hand,…
  - DNA mapping
- Dynamic pattern
  - Signature, voiceprint,…
- Applications
  - Access control
  - Electronic transaction authentication, etc.




www.research.ibm.com



## Gesture Recognition

- Sign language interpretation
- Object manipulation in Virtual Reality

www.5dt.com

- Tele-operations
  - Control remote by gesture input
  - TV control by hand motion

Photo credit: Ben-Gurion University of the Negev, Israel

## Optical Character Recognition

- Reconstructing text from printed materials (e.g. reading machines)
- License plate identification
- Sorting mails by recognizing barcode, postal code, or address


http://www.plate-recognition.info

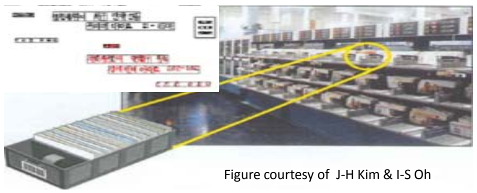


Figure courtesy of J-H Kim & I-S Oh



## Handwritten Character Recognition

- Input devices
  - Mobile devices with camera or scanner
  - Motion sensing pen, etc.
- Example: Freehand equation input

$$\sum_{y=1}^{m}\sum_{x=1}^{n}\frac{2xy}{1+x^3} \qquad \int_{x=1}^{\infty}\frac{f(x)}{e^{-x}}dx$$

Equation input with electronic pen

$$\sum_{y=1}^{m}\sum_{x=1}^{n}\frac{2xy}{1+x^3} \qquad \int_{x=1}^{\infty}\frac{f(x)}{e^{-x}}dx$$

Equation recognition



## Electronic Field Guide & AR

- Leafsnap (http://leafsnap.com)

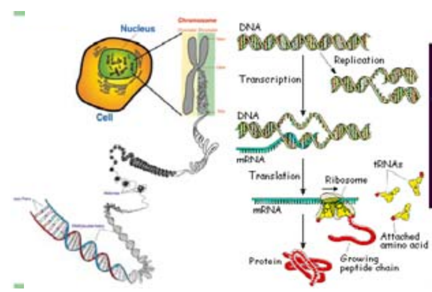

iPhone screenshot from http://itunes.apple.com/

- Google glasses




## Computational Biology

- DNA sequence identification
- Protein structure analysis



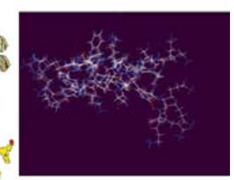



Figure courtesy of J-H Kim

# Data Mining

- Extract patterns from data

| Data | → | Information | → | Decision |
|---|---|---|---|---|
| • Census<br>• Point of sale<br>• ATM<br>• Banking statistics<br>• Credit information<br>• Record of medical examination | | • 80% of the customers who purchased A also buy B<br>• Purchasing power in US market has been decreased during the last 6 months<br>• Sales of A increased twice than B<br>• It's dangerous if patients show dehydration | | • Marketing strategy?<br>• Display of goods<br>• Optimal budget assignment?<br>• Plan to increase market share?<br>• Prescription? |

- Example: Prevention of lost/stolen credit card by analyzing patterns of card usage

Slide courtesy of J-H Kim



# And More…

- Prediction system
  - Weather forecasting based on the satellite data
- Security and military applications
  - computer security: identify threats through analysis of network traffic patterns
  - Aerial reconnaissance (e.g. recognizing missile silos, airfields, etc.)
  - Radar signal classification
- Dating services
  - Pattern includes age, sex, hobbies, income, etc.



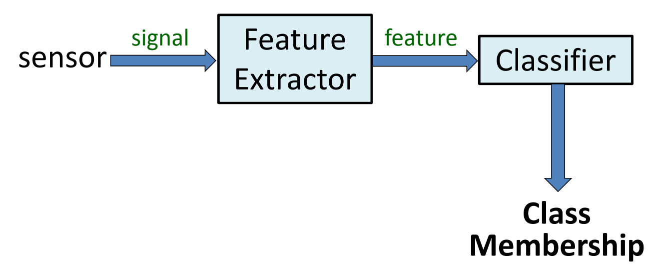
# Terminology

- **Pattern:** object, process or event consisting of both deterministic and stochastic components; a record of dynamic occurrences influenced by both deterministic and stochastic factors
- **Pattern Class:** set of patterns sharing a set of common attributes (or features) and usually originating from the same source (associated with the *generalization* or *abstraction* of patterns)
- **Features:** relevant (intrinsic) trait or characteristic that makes a pattern apart from another; data extractable through measurement and/or processing
- Examples
  - **Patterns:** textures, biology patterns, constellations, weather pattern, speech waveform, etc.
  - **Features:** color, age, weight, aspect ratio, etc.
- **Noise:** distortions associated with pattern processing (errors in feature extraction) and/or training samples that impact the classification abilities of the system

Slide courtesy of S. Iliescu

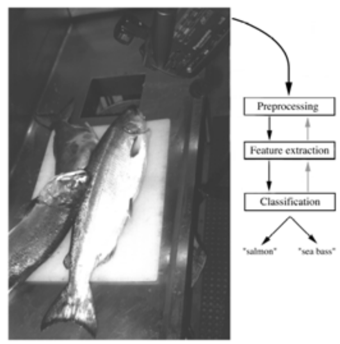


# Simple Process of PR



# Classification Example

- Sorting incoming fish on a conveyor belt into two species (Sea bass or Salmon) using optical sensing
- Preprocessor
  - Simplify subsequent operations without losing relevant information
  - Segmentation: isolate fish
- Feature extractor
  - Reduce the data by measuring certain *features* or *properties*
- Classifier
  - Evaluates the evidence presented and makes a final decision as to the species





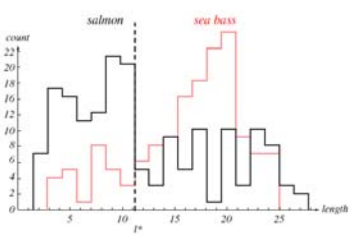

# Salmon vs Sea Bass Discrimination

- Features
  - Physical differences between the two types of fish
    - Length, lightness, width, number and shape of fins, position of the mouth, etc…
  - Noise or variations in the images
    - Variations in lighting, position of the fish
- Models
  - Differences between the population of sea bass and that of salmon
  - Different descriptions in mathematical form
  - Suppose somebody at the plant tell us that a sea bass is generally longer than a salmon:
    - Tentative model: *sea bass have some typical length and this is greater than that for salmon*

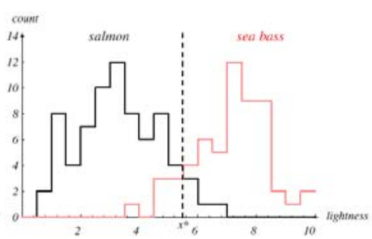


# Salmon vs Sea Bass

- A single feature cannot discriminate




Histograms for the length feature




Histograms for the lightness feature



# Cost of Misclassification

- There are two possible classification errors
  - Deciding a sea bass into a salmon
  - Deciding a salmon into a sea bass
- Which error is more important?
- Generalized as "cost function" (or "loss function")
- Then, look for the decision of minimum Risk
  - Risk = Expected Loss

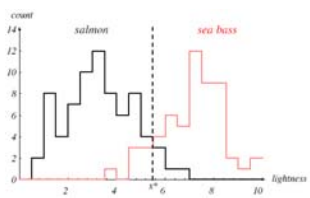
| truth \ decision | Salmon | Sea Bass |
|---|---|---|
| Salmon | 0 | –10 |
| Sea Bass | –20 | 0 |

Cost Function

# Salmon vs Sea Bass

- Threshold decision boundary and cost relationship
  - Move our decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of sea bass that are classified as salmon!)

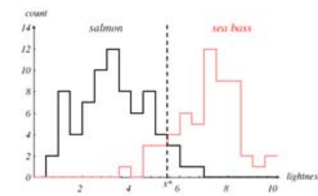
Central task of Decision Theory






# Decision Theory

- Make a decision rule (i.e. set a decision boundary) so as to minimize a cost
- A classifier partitions the feature space into class-labeled decision regions. The border of each region is a *decision boundary*
- Ideally, a unique class assignment is achieved if, the decision regions cover the entire feature space and they are disjoint (i.e. the decision boundaries do not create error or confusion areas)

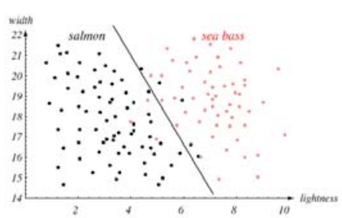



partly from S. Iliescu



# Back to the Fish Example

- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such “noisy features”
- Adopt the lightness and add the width of the fish




Fish $\longrightarrow \mathbf{x}^T = [x_1, x_2]$

Lightness Width

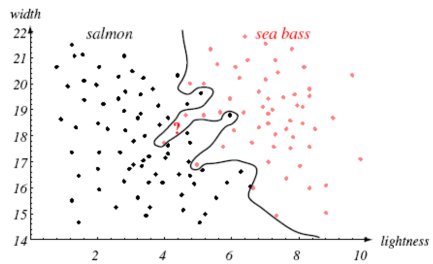


# How many features and which?

- Choice of features determines success or failure of classification task
- For a given feature, we may compute the best decision strategy from the (training) data
  - Is called training, parameter adaptation, learning
  - Machine learning issues
- Issue with feature extraction
  - Correlated features do not improve performance
  - It might be difficult to extract certain features
  - It might be computationally expensive to extract many features
  - Curse of dimensionality

# Overfitting

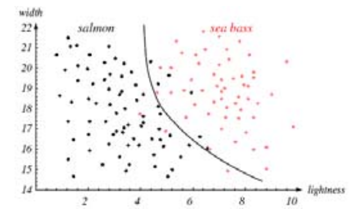
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:






# Generalization

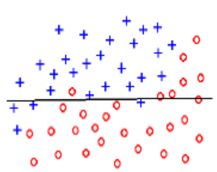
- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify *novel* input ⟶ Issue of *generalization*!
- How can we improve generalization performance?
  - More training examples better estimates of the true underlying characteristics
  - Simple models (i.e. simpler classification boundaries) *usually* yield better performances



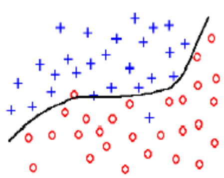

Simplify the decision boundary!



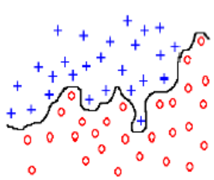
# Overfitting and Underfitting



underfitting



good fit



overfitting

Figures courtesy of J-H Kim



# PR/Classification Strategy

- Extract the feature vector of the pattern
- Compare the feature vector with the characteristics of existing decision regions (i.e. with features of each pattern class)
- Assign pattern to the class (associated with decision region) matching the extracted feature vector
- Basic ingredients
  - Measurement space (e.g. image intensity)
  - Features (e.g. edges/corners, spectral energy)
  - Classifier: soft or hard
  - Decision boundary
  - Training samples

## PR Approaches

- *Basic assumption*: there is an underlying model behind the observed phenomena
- *Question*: based on noisy observations, what is the underlying model?
- **Template matching:** the pattern to be recognized is matched against a stored template
  - Works only for simple problems
- **Statistical**: based on underlying statistical model of patterns (features) and pattern classes
  - Class-conditional probability $p(\mathbf{x}|C_i)$: usually learned from examples



## PR Approaches (cont'd)

- **Structural**: assume that interrelations between parts (features) are more important, and finding them (though not easy) may require prior knowledge about internal structure
  - Pattern classes represented by means of formal structure as rules, grammars, automata, strings, etc.
  - Based on measures of structural similarity
  - Often called *syntactic* pattern recognition
- **Neural network**: classifier is represented as a network of cells modeling neurons of the human brain (connectionist approach)
  - Knowledge is stored in the connectivity and strength of synaptic weights
  - Trainable, requires minimum a priori knowledge, works for complex decision boundaries



## Example: Neural, Statistical, and Structural OCR

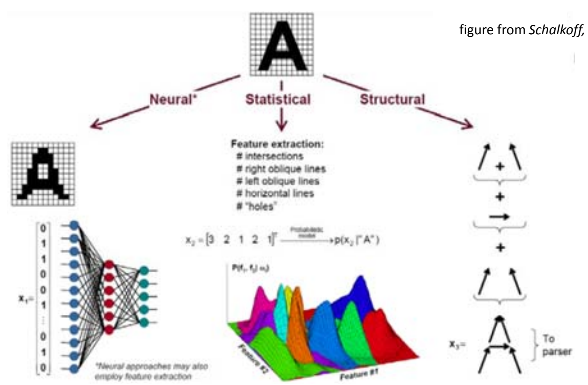
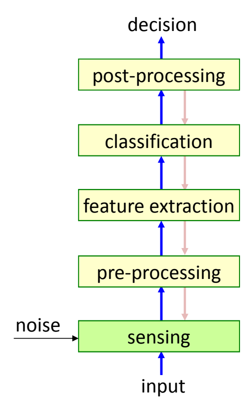



figure from *Schalkoff, 1992*

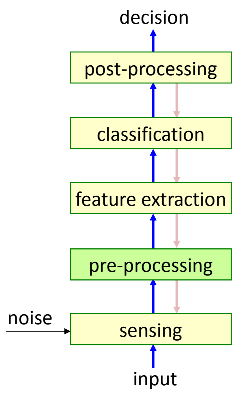


## Pattern Recognition Systems




- Sensing: data collection
  - Input pattern is received via a sensor, a transducer, etc. (e.g. camera or microphone)
  - PR system depends on the transducer's bandwidth, resolution, sensitivity, distortion, signal-to-noise ratio, etc.

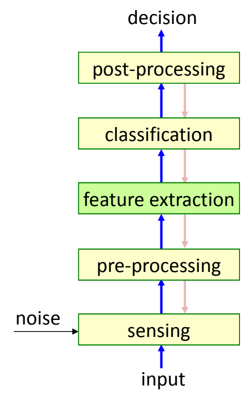
## Pattern Recognition Systems




- Preprocessing
  - Segmentation and grouping
    - Patterns should be well separated and should not overlap
    - One of the deepest problems in PR
  - Noise filtering
  - Size normalization
  - Etc…



## Pattern Recognition Systems

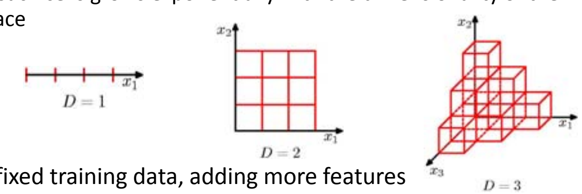



- Feature extraction
  - Discriminative features
  - Invariant features with respect to irrelevant transformations (e.g. translation, rotation and scale) and other variations
  - Feature selection
    - Select the most valuable features from a larger set of candidates (identifying pattern traits)



## Curse of Dimensionality

- Adding too many features can, paradoxically, lead to a worsening of performance
  - If the feature space is divided into regular cells, then the number of such cells grows exponentially with the dimensionality of the space

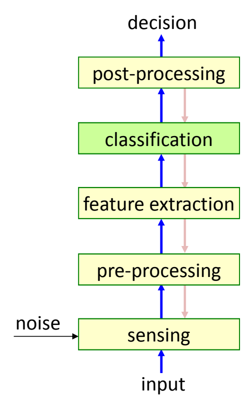



- With fixed training data, adding more features worsening the quality of generalization
  - Too many cell causes small number of training samples per cell → poor generalization

Figures from *PRML* by C. Bishop

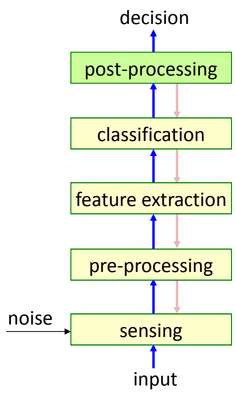


## Pattern Recognition Systems




- Classification
  - Applying recognition scheme
  - Match pattern to class: use a feature vector provided by a feature extractor to assign the object to a category
  - Adjust for missing feature

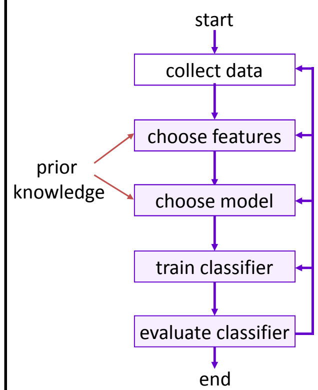
# Pattern Recognition Systems




- Post Processing
  – Decide on the recommended action using the output of the classifier
  – Exploit context – input dependent information other than from the target pattern itself – to improve performance
  – Adjustments for context (apply correction techniques)



# Pattern Recognition Design Cycle




- Data collection
  – Probably the most time-intensive component of project
  – How many examples are enough?
- Feature choice
  – Critical to the success of the PR project
  – Require basic prior knowledge
- Model choice and design
  – Statistical, neural and structural?
  – Parameter settings
- Training
  – Given a feature set and 'blank' model, adapt the model to explain the training data
  – Supervised, unsupervised, reinforcement learning
- Evaluation
  – How well does the trained model do?
  – Overfitting vs. generalization