

In The Name of Allah



Digital Media Laboratory
Sharif University of Technology

Statistical Pattern Recognition

A Brief Overview of the course

Hamid R. Rabiee

Jafar Muhammadi, Nima Pourdamghani

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

Agenda

- ✧ **What is a Pattern?**
- ✧ **What is Pattern Recognition (PR)?**
- ✧ **Applications of PR**
- ✧ **Components of a PR system**
- ✧ **Features**
- ✧ **Types of Learning**
- ✧ **The Design Cycle**
- ✧ **Pattern Recognition Approaches**
- ✧ **Brief Mathematical Overview**
- ✧ **Course Road Map**



What is a pattern?

✧ **Pattern**

✧ **Opposite to chaos; it is an entity, object, process or event, vaguely defined, that can be given a name or “label”.**

✧ **For example, a pattern could be**

✧ **A fingerprint image**

✧ **A handwritten cursive word**

✧ **A human face**

✧ **A speech signal**

✧ **Texture**

✧ **Etc.**



What is Pattern Recognition?

✧ **Pattern recognition (PR)**

- ✧ **The study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns.**
- ✧ **The assignment of a physical object or event to one of several pre-specified categories**

✧ **Related terms**

- ✧ **A *pattern class* (or category) is a set of patterns sharing common attributes and usually originating from the same source.**
- ✧ **During *recognition* (or *classification*) given objects are assigned to prescribed classes (get labeled).**
- ✧ **A *classifier* is a machine which performs classification.**



An Example

- ✧ **Four pattern categories (classes)**

- ✧ **Sea, Beach, Jungle, Sky**

- ✧ **Common attributes (features)**

- ✧ **Color**

- ✧ **Contrast**

- ✧ **Texture**

- ✧ **Goal**

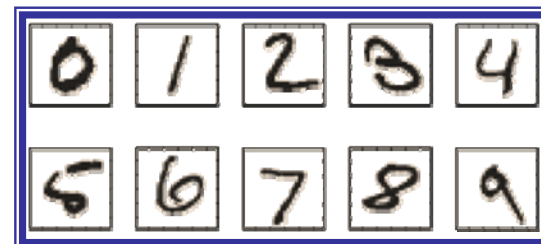
- ✧ **Observing some labeled pixels, We wish to assign a label to each new (unlabeled) pixel.**



Applications of PR

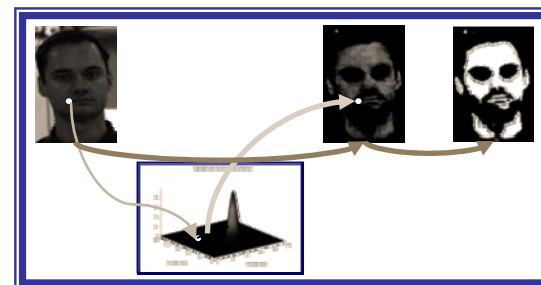
✧ Handwritten Digit Recognition

- ✧ Input Pattern: pictures of handwritten digits
- ✧ Output Classes: the digit (0 .. 9)



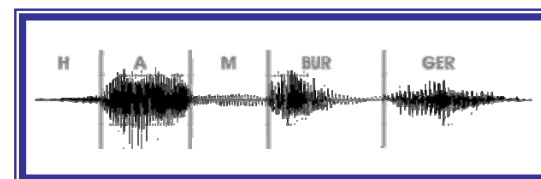
✧ Skin Detection

- ✧ Input pattern: a picture
- ✧ Output Classes: Skin / not skin for each pixel



✧ Speech Recognition

- ✧ Input Pattern: Speech waveform
- ✧ Output Classes: Specified Spoken words



Applications of PR

- ✧ **Document Classification (Web news classification)**
 - ✧ **Input Pattern:** Text or html document
 - ✧ **Output Classes:** Semantic Categories (e.g. business, sports, ...)

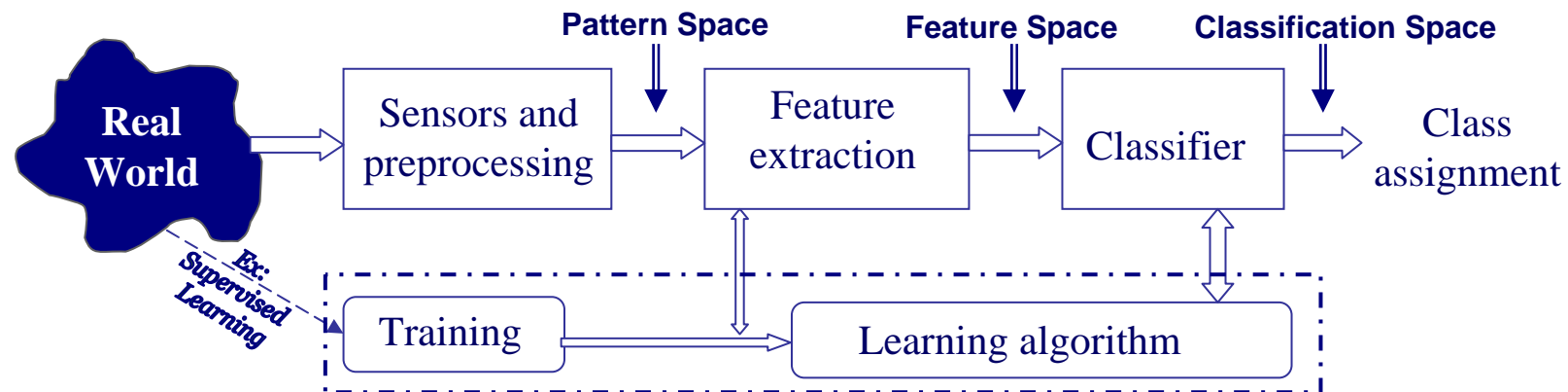
- ✧ **Financial Time Series Prediction**
 - ✧ **Input Pattern:** relation between consecutive data of time series
 - ✧ **Output Values:** possible values of output (regression problem)

- ✧ **Sequence Analysis (Bioinformatics)**
 - ✧ **Input Pattern:** DNA / protein sequences
 - ✧ **Output:** Known types of genes

- ✧ **Spam Detection**
 - ✧ **Input Pattern:** Text / image of emails
 - ✧ **Output Classes:** Spam / not spam



Components of a PR system



✧ Components

- ✧ **Sensors and preprocessing**
- ✧ **feature extraction**
- ✧ **Classifier**
- ✧ **Training: Provides some useful information for supervised learning**
- ✧ **Learning algorithm: Create classifier from training data (labeled samples)**



Components of a PR system

✧ Example: Separate different types of fishes

✧ Sensor:

- ✧ Camera

✧ Preprocessing:

- ✧ Segmentation

✧ Features:

- ✧ Ask experts the major differences between types
- ✧ See different fishes and find the differences
- ✧ Length, Width, Number of fins, ...

✧ Learning:

- ✧ Ask experts the type of sample fishes
- ✧ Find typical length of each type

✧ Classification:

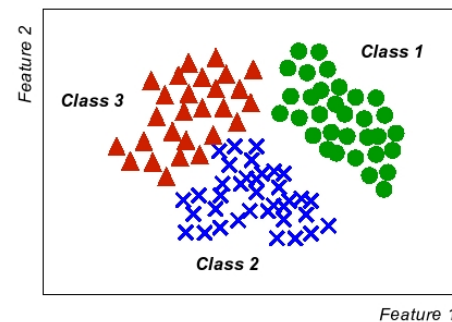
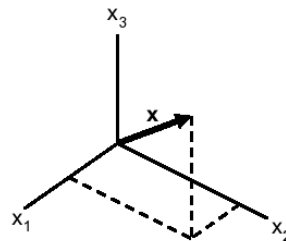
- ✧ Compare the length (width, etc) of a new fish to the learned lengths



Features

- ✧ **Feature is any distinctive aspect, quality or characteristic**
 - ✧ Features may be symbolic (i.e., color) or numeric (i.e, height)
- ✧ **Definitions**
 - ✧ The combination of d feature is represented as a d-dimensional column vector called a feature vector.
 - ✧ The d-dimensional space defined by the feature vector is called the feature space.
 - ✧ Objects are represented as points in feature space. This representation is called a scatter plot.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_d \end{bmatrix}$$



Feature vector

Feature space

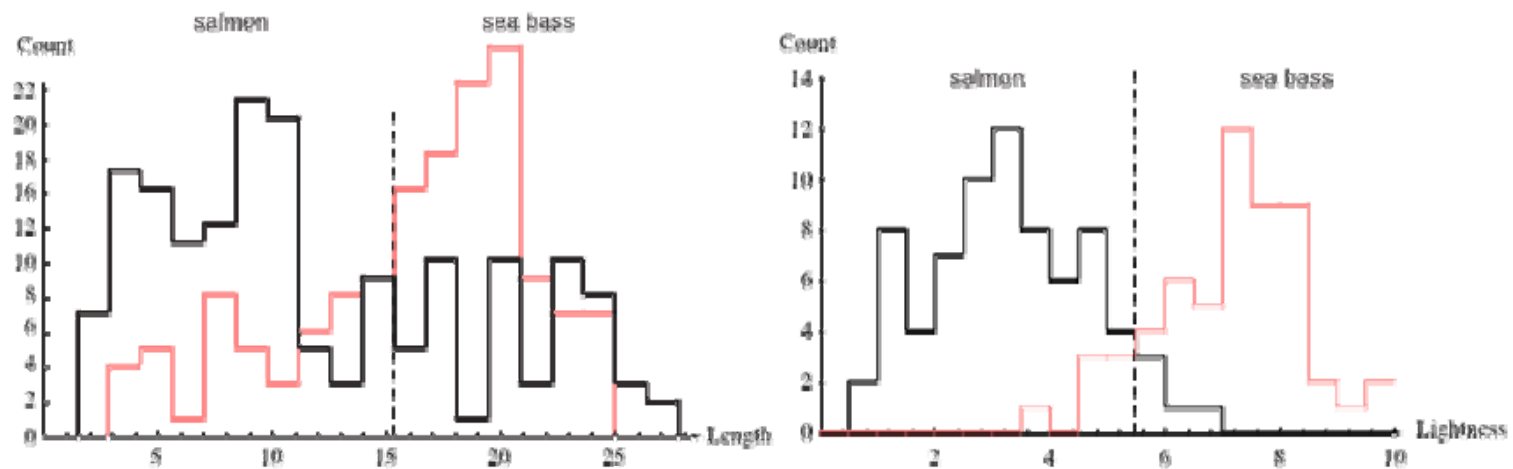
Scatter plot



Features

❖ “Fish Separation” Example:

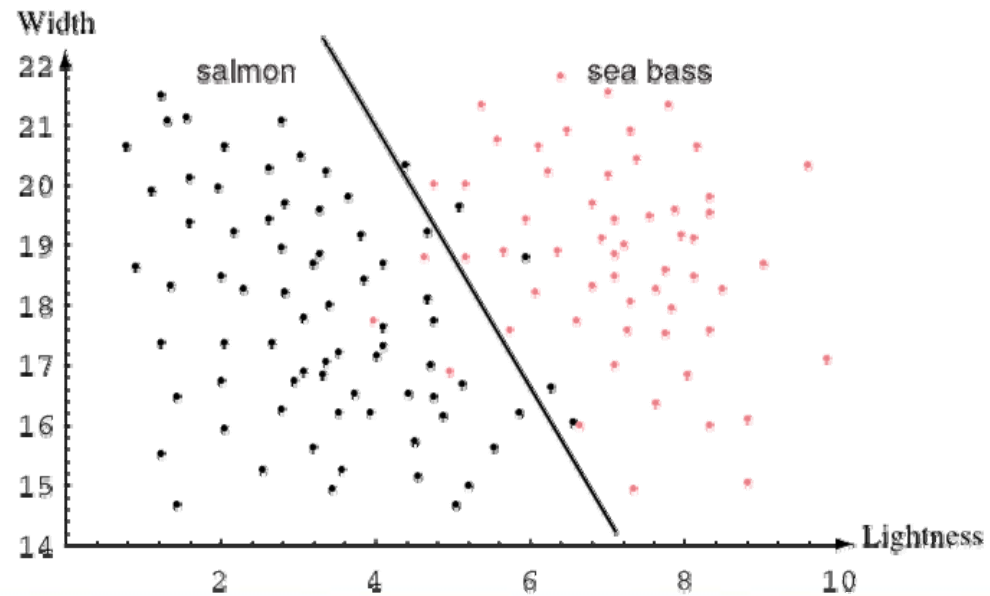
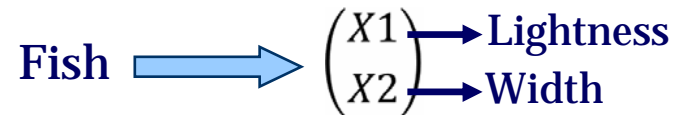
- ❖ The length is a poor feature alone!
- ❖ Select the lightness as a possible feature



Features

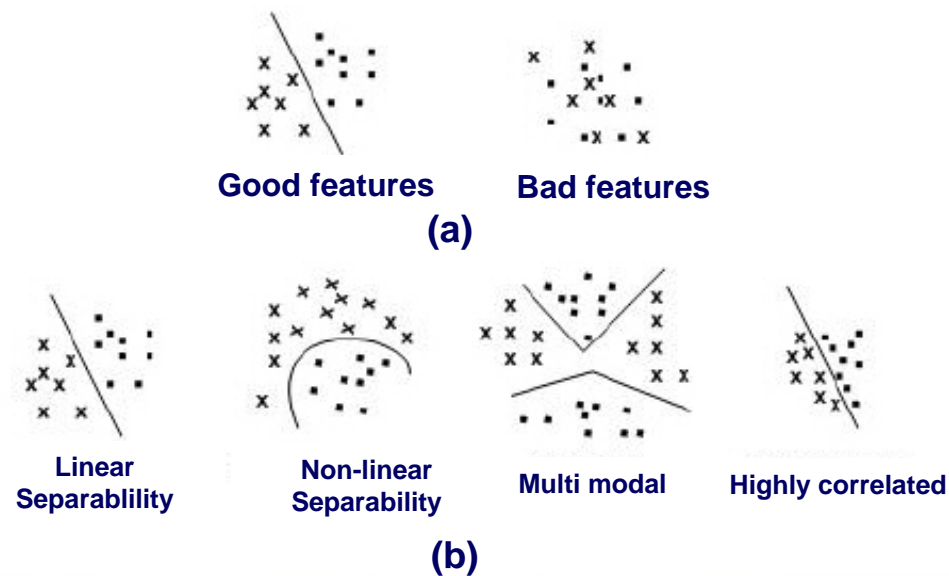
✧ “Fish Separation” Example:

✧ The scatter plot for following two features:



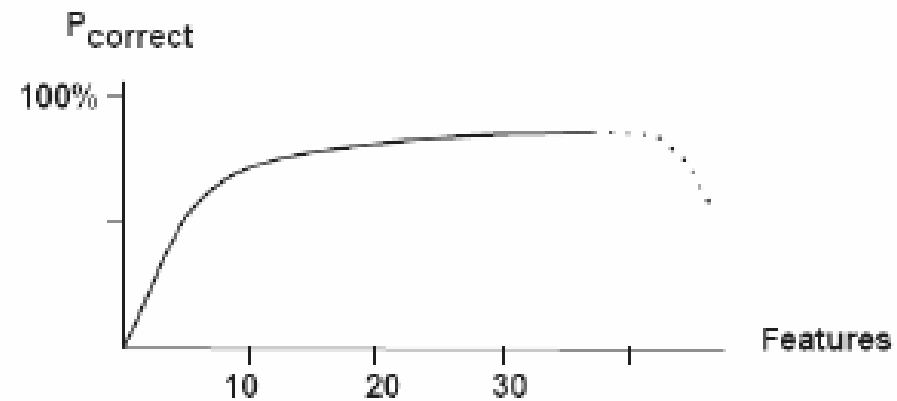
Good/Bad Features & Classification

- ✧ The quality of a feature vector is related to its ability to discriminate examples from different classes
 - Examples from the same class should have similar feature values
 - Examples from the different classes have different feature values
- ✧ The distinction between good and poor features (a), and feature properties (b)



Feature Dimension

- ✧ **The curse of dimensionality**
- ✧ **The probability of misclassification of a decision rule does not decrease beyond a certain dimension for the feature space as the number of features increases.**

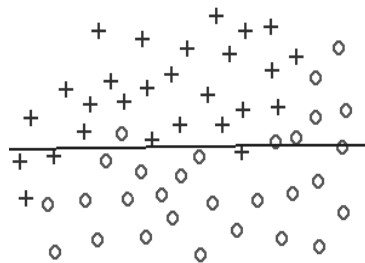


- ✧ **Peaking Phenomena**
 - ✧ **Adding features may actually degrade the performance of a classifier**

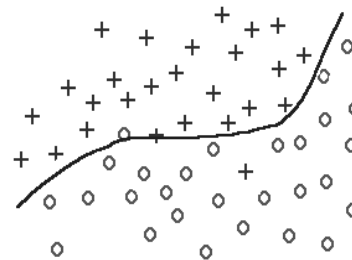


Overfitting and underfitting

- ✧ The inductive bias of a learning algorithm is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered (Mitchell, 1980)
- ✧ Too high a bias, though simpler, may lead to underfitting and too low a bias, on the other hand, evidently more complex, may lead to overfitting.



underfitting



good fit

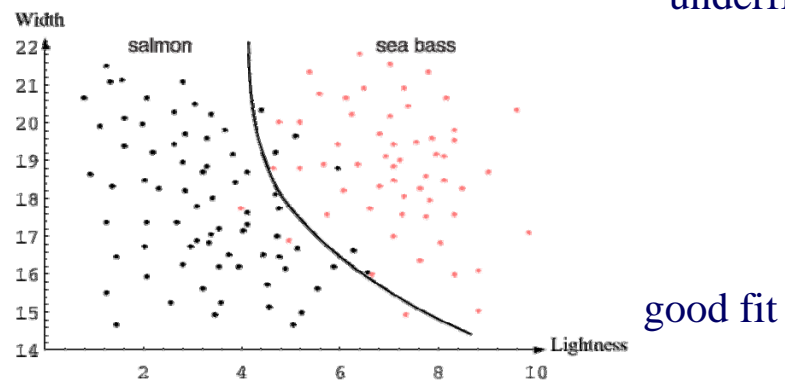
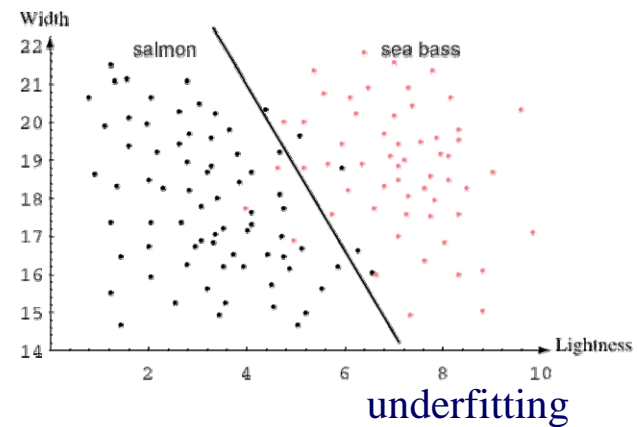
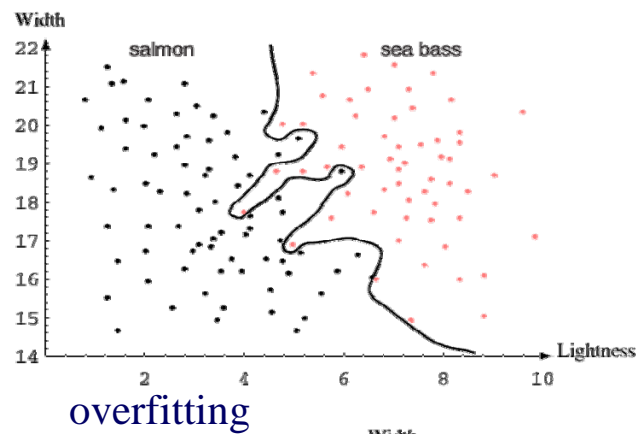


overfitting



Overfitting and underfitting

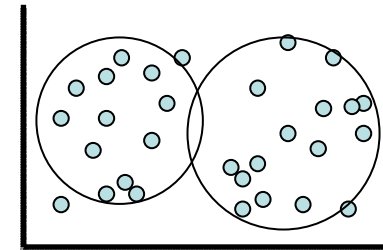
✧ “Fish Separation” Example:



Types of Learning

✧ Unsupervised learning

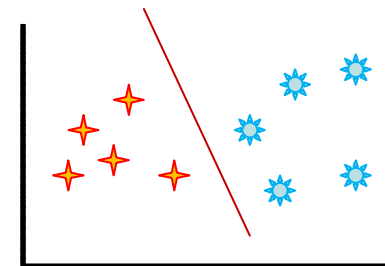
- ✧ The system forms clusters or “natural groupings” of the input patterns



Unsupervised learning

✧ Supervised learning

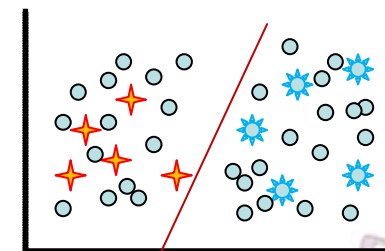
- ✧ Classify data using labeled samples (labels are provided by a trainer)



Supervised learning

✧ Semi-supervised learning

- ✧ make use of both labeled and unlabeled data for training - typically a small amount of labeled data with a large amount of unlabeled data.



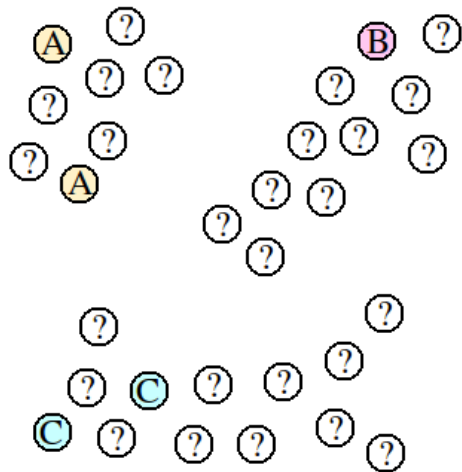
Semi-Supervised learning



Types of Learning (Algorithmic view point)

✧ Types of Learning

- ✧ **Inductive: Learns a labeling function over the space**
- ✧ **Transductive: Just labels the given test queries**



From Wikipedia

- The goal is to predict appropriate labels for all of the unlabeled points shown as (?).

- The **Inductive** learning algorithm will only have five labeled points to use as a basis for building a predictive model. For example, if a nearest-neighbor algorithm is used, then the points near the middle will be labeled "A" or "C" instead of "B".

- **Transduction** has the advantage of being able to consider all of the points, not just the labeled points, while performing the labeling task. In this case, transductive algorithms would label the unlabeled points according to the clusters to which they naturally belong. The points in the middle, therefore, would most likely be labeled "B",

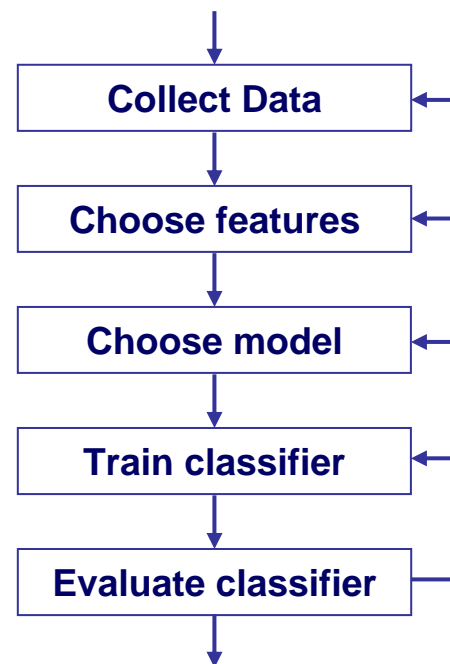
- One disadvantage of transduction is that it builds no predictive model. If a previously unknown point is added to the set, the entire transductive algorithm would need to be repeated with all of the points in order to predict a label.



The Design Cycle

✧ Design cycle

- ✧ Data collection
- ✧ Feature Choice
- ✧ Model Choice
- ✧ Training
- ✧ Evaluation
- ✧ Computational Complexity



More on Design Cycle

✧ **Data Collection**

- ✧ How much data is sufficient and representative?

✧ **Feature Choice**

- ✧ **Domain Specific**
- ✧ **Good features:**
 - ✧ Simple to extract, invariant to irrelevant transformation, insensitive to noise.

✧ **Model Choice**

- ✧ Which model to choose for better performance?

✧ **Training**

- ✧ Which of the many different procedures?

✧ **Evaluation**

- ✧ Measure the error rate

✧ **Computational Complexity**

- ✧ What is the trade-off between computational ease and performance?



Pattern Recognition Approaches

✧ **Statistical PR**

- ✧ **Based on underlying statistical properties/model of patterns and pattern classes.**
- ✧ **Use numerical features for distinguishing between classes.**
 - ✧ Bayesian Methods
 - ✧ Neural Networks
 - ✧ Decision Trees
 - ✧ Support Vector Machines
 - ✧ Etc.

✧ **Structural (or syntactic) PR**

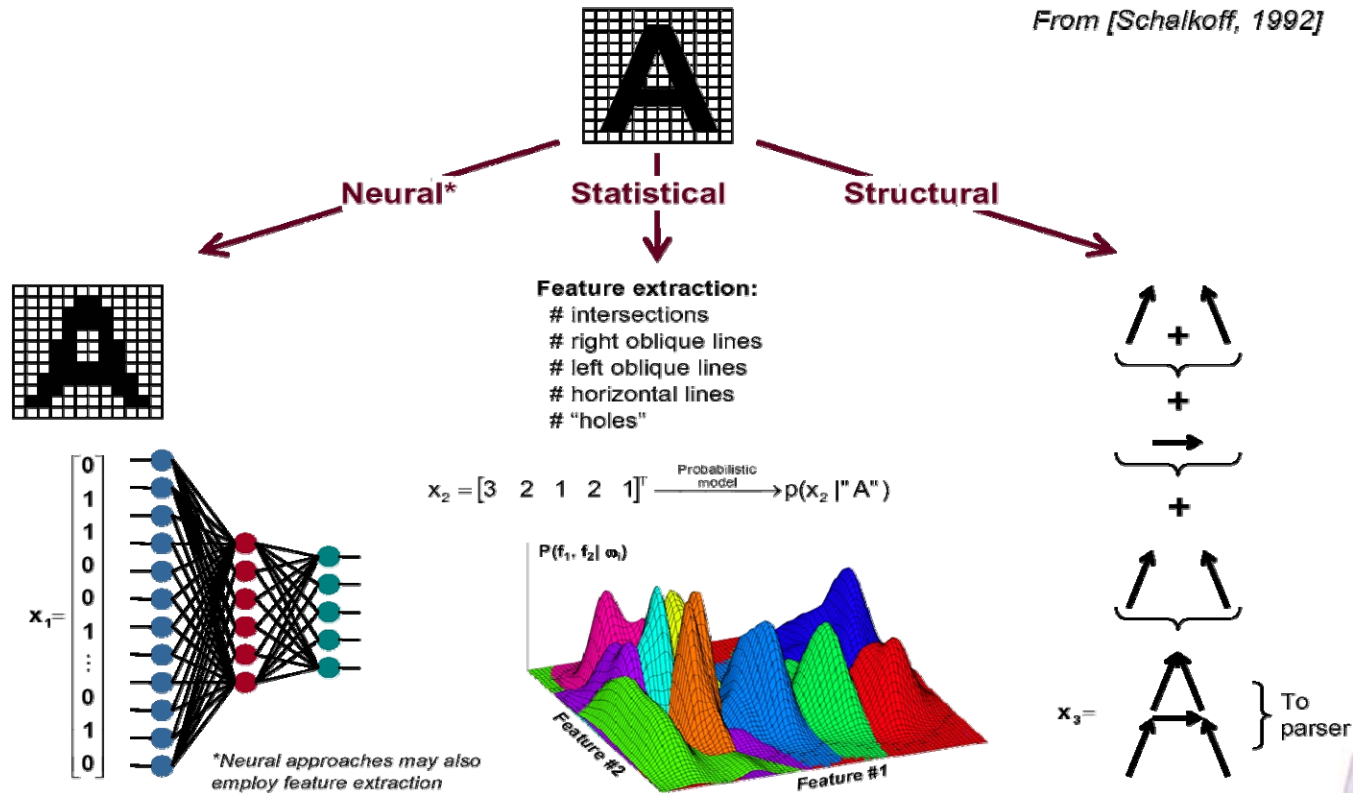
- ✧ **Based on explicit or implicit representation of a class's structure**
- ✧ **Pattern classes represented by means of formal structures as grammars, automata, strings, graphs, trees, etc.**
- ✧ **Reference: “Syntactic and structural pattern recognition: theory and applications”, By Horst Bunke, Alberto Sanfeliu**



Pattern Recognition Approaches

Example: Example: neural, statistical and structural OCR

From [Schalkoff, 1992]



Template designed by Jafar Muhammad



Background Mathematical Review

- ✧ **In the TA class (attendance mandatory) you'll review the following mathematical concepts:**
 - ✧ **Distribution functions and measures**
 - ✧ Distribution functions
 - ✧ Moments, Covariance matrix
 - ✧ **Feature spaces (correlation, orthogonality, independency, etc.)**
 - ✧ **Gaussian Distribution**
 - ✧ Gaussian distribution function
 - ✧ Central Limit Theorem
 - ✧ **Linear Algebra**
 - ✧ Matrices (rank, determinant, inversion, differential, derivation, etc)
 - ✧ Eigen values and eigen vectors
 - ✧ **Information theory**
 - ✧ Entropy, Information gain, etc.
 - ✧ **Distances**
 - ✧ Axioms of distance measure
 - ✧ Distance measures: Euclidean, Mahalanobis, Minkowsky, etc.
 - ✧ Distributions distance measure: Kullback-Leibler



Road Map

- ✧ **How to choose / extract features**
- ✧ **Dimensionality reduction**
- ✧ **Classification**
 - ✧ Probabilistic methods
 - ✧ Linear discriminant methods
 - ✧ Non parametric methods
 - ✧ Neural Networks
 - ✧ Support Vector Machines
 - ✧ Kernel methods
 - ✧ Graphical methods
- ✧ **Clustering**
 - ✧ Partitioning methods
 - ✧ Density based methods
 - ✧ Expectation Maximization
- ✧ **Semi-supervised learning**
- ✧ **Applications**



Any Question

End of Lecture 1

Thank you!

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

