

Pattern Recognition: Theory and Applications

Saroj K. Meher

**Systems Science and Informatics Unit,
Indian Statistical Institute, Bangalore Centre.**

Contents:

- Pattern recognition Introduction
- Definitions and framework
- Advantages and Applications areas
- Discrimination functions
- Learning methodologies
- Some examples of classification and Clustering methods
- Semisupervised and co-training algorithm
- Multiple Classification Systems
- Contd...

Pattern Recognition (PR): Motivation

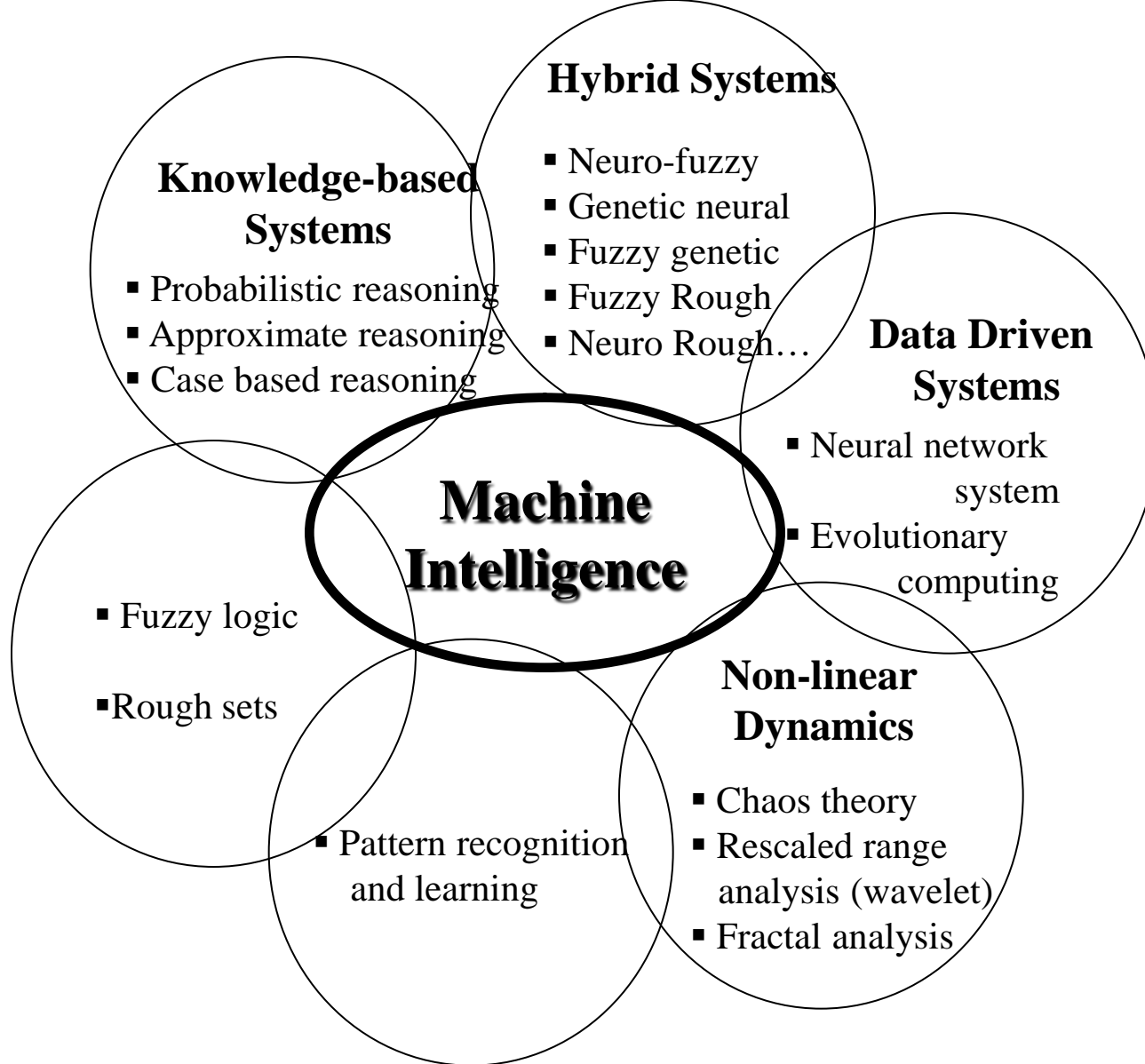
- Pattern recognition stems from the need for automated machine recognition of objects, signals or images, or the need for automated decision-making based on a given set of parameters.

PR Defined

PR is a domain of machine learning. It is the act of taking in raw data and make an action based on it.

- The assignment of a physical object or event to one of several pre-specified categories” –*Duda and Hart*
- A problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories or classes” – *Fukunaga*
- The science that concerns the description or classification (recognition) of measurements” –*Schalkoff*
- The process of giving names ω to observations x ”, –*Schürmann*
- PR is concerned with answering the question “*What is this?*” –*Morse*

PR PLAYS THE ROLE WHEN A DECISION IS ABOUT TO TAKE



Machine Intelligence: A core concept for grouping various advanced technologies with **Pattern Recognition and Learning**

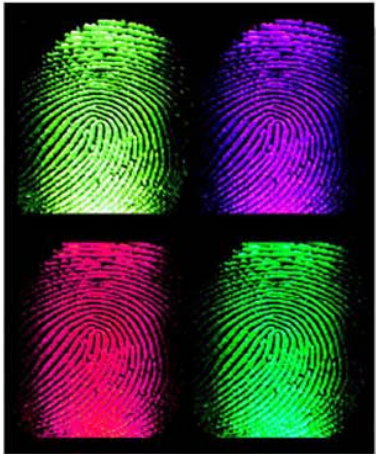
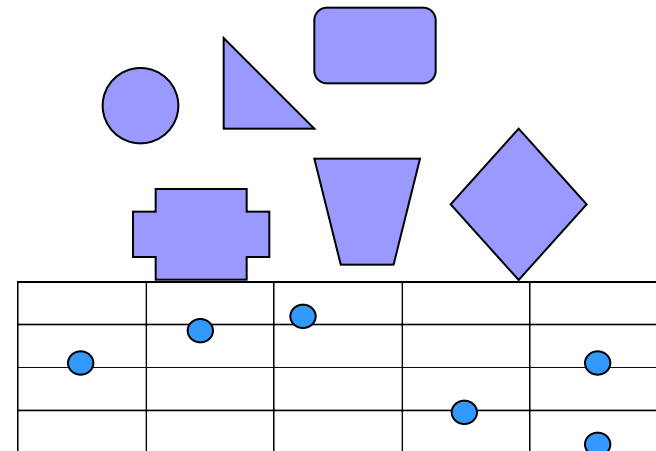
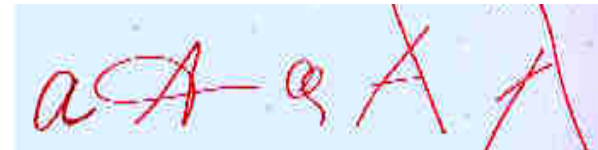
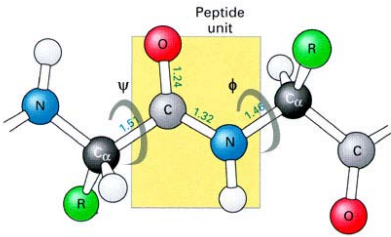
Pattern Recognition System (PRS)

Measurement → *Feature* → *Decision*
Space *Space* *Space*

What is a pattern?

“A pattern is the opposite of a chaos; it is an entity vaguely defined, that could be given a name.”

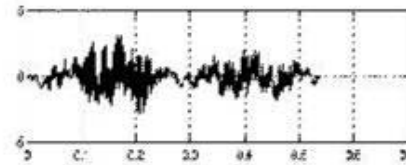
WHAT ABOUT TEXTURE ?



2012

What is a pattern?

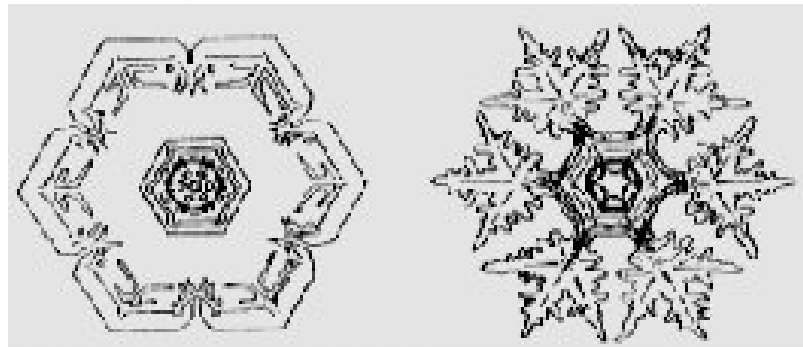
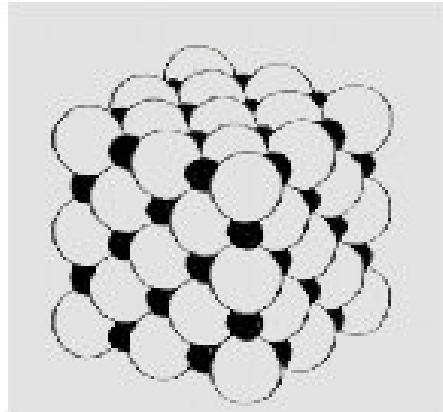
A pattern is an abstract object, or a set of measurements describing a physical object.



John Smith

Examples of Patterns

Cristal Patterns: atómico or molecular



Their structures are represented by 3D graphs and can be described by deterministic grammars or formal languages

Examples of Patterns

Patterns of Constellations

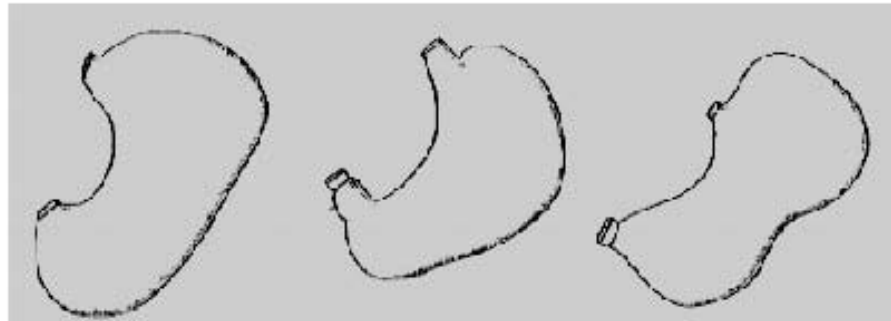
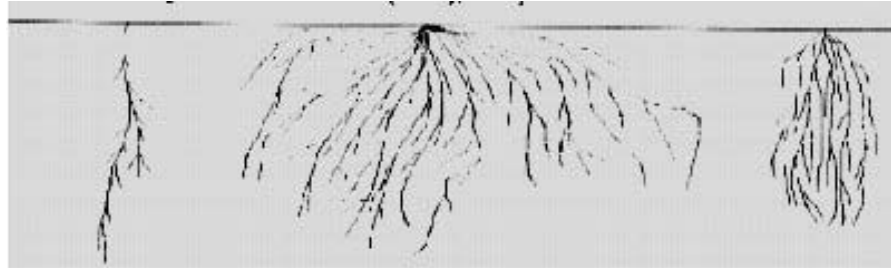
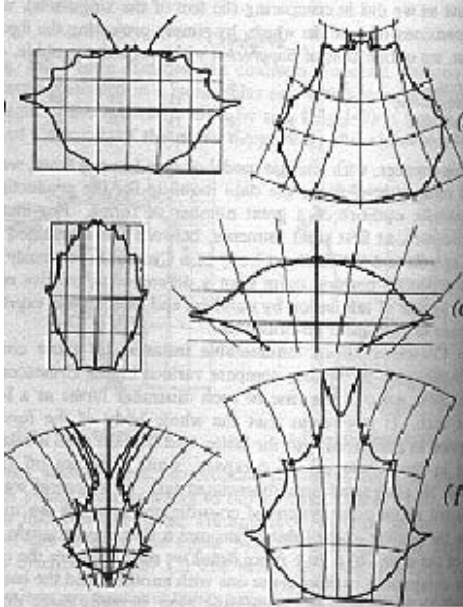


Patterns of constellations are represented by 2D planar graphs

Human perception has strong tendency to find patterns from anything. We see patterns from even random noise --- we are more likely to believe a hidden pattern than denying it when the risk (reward) for missing (discovering) a pattern is often high.

Examples of Patterns

Biological Patterns ---

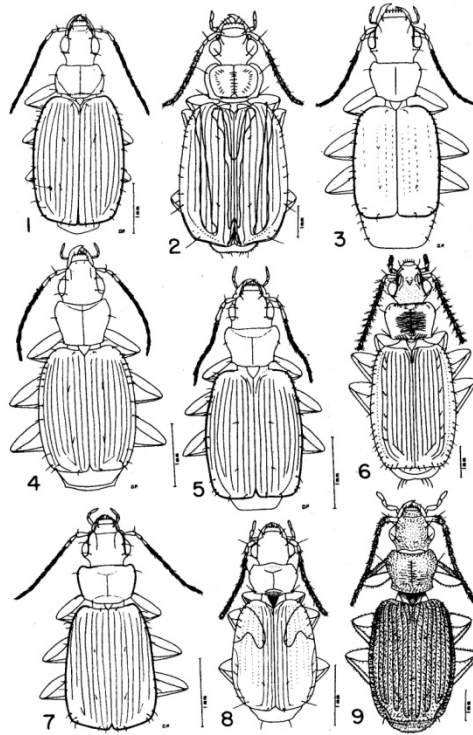


Landmarks are identified from biologic forms and these patterns are then represented by a list of points. But for other forms, like the root of plants, Points cannot be registered crossing instances.

Applications: Biometrics, computacional anatomy, brain mapping, ...

Examples of Patterns

Biological Patterns



Landmarks are identified from biologic forms and these patterns are then represented by a list of points.

Examples of Patterns

Music Patterns

51

The image displays five systems of musical notation for piano, each featuring a circled pattern. The systems are numbered 318, 323, 328, 332, and 336. The notation includes treble and bass clefs, notes, rests, and dynamic markings such as *pp*, *f*, *dn.*, and *mf*. Fingerings are indicated by numbers 1-5. The patterns are circled in black, highlighting specific melodic or rhythmic motifs. The first system (318) shows a pattern in the right hand starting with a circled note. The second system (323) has a pattern in the right hand. The third system (328) has a pattern in the right hand. The fourth system (332) has a pattern in the right hand. The fifth system (336) has a pattern in the right hand.

Ravel Symphony?

Examples of Patterns

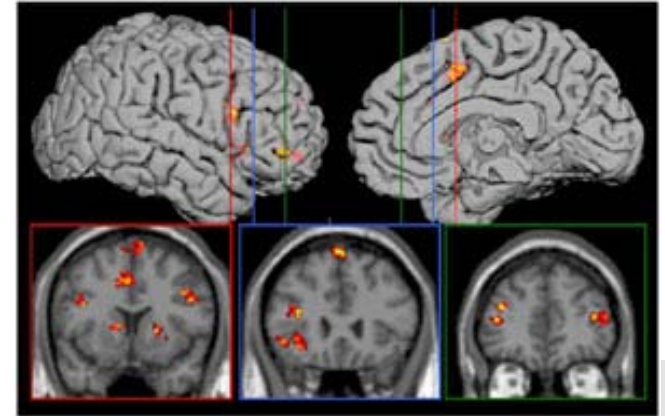
People Recognition



Funny, Funny

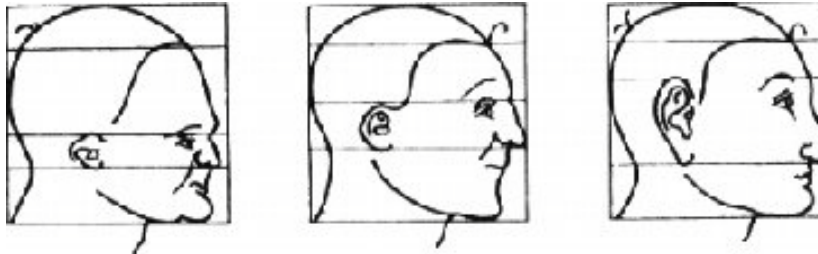


Patterns Behavior ?



Examples of Patterns

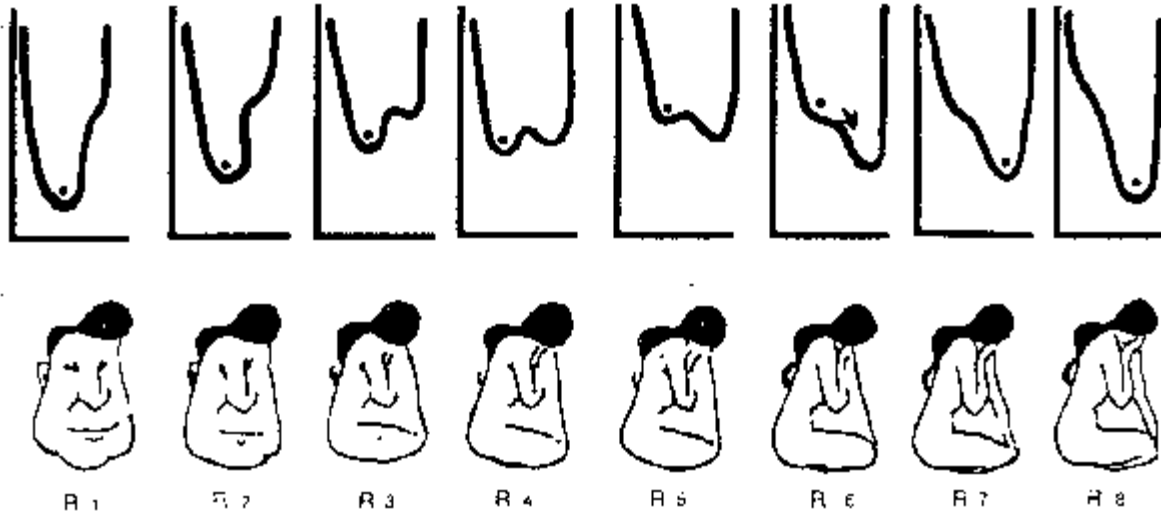
Discovery and Association of Patterns



Statistics show connections between the shape of one's face (adults) and his/her Character. There is also evidence that the outline of children's face is related to alcohol abuse during pregnancy.

Examples of Patterns

Discovery and Association of Patterns

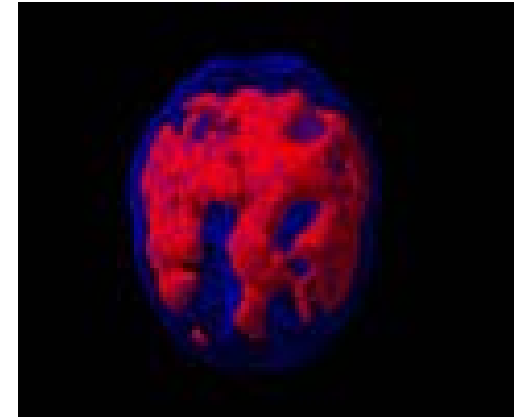
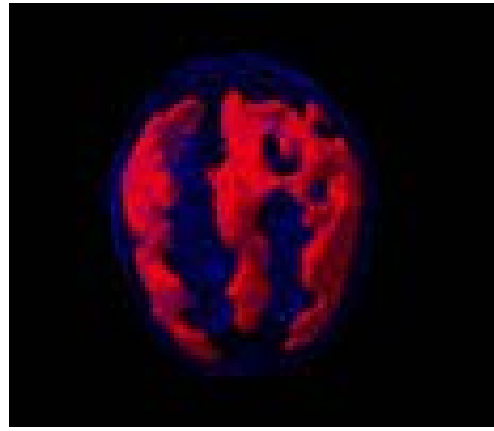
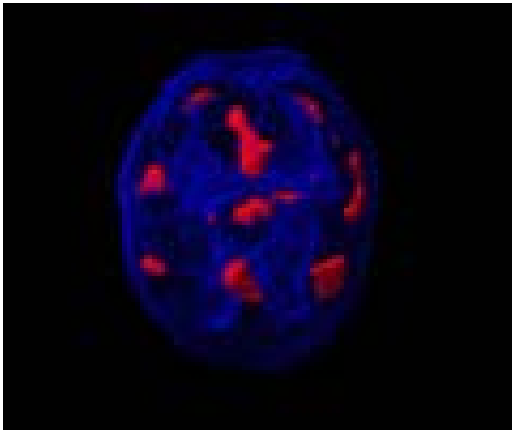


What are the features?

Statistics show connections between the shape of one's face (adults) and his/her Character.

Examples of Patterns

Patterns of Brain Activity



We may understand patterns of brain activity and find relationships between brain activities, cognition, and behaviors

Examples of Patterns

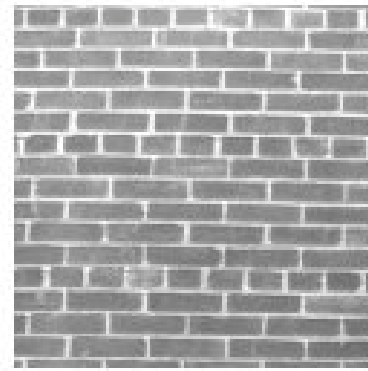
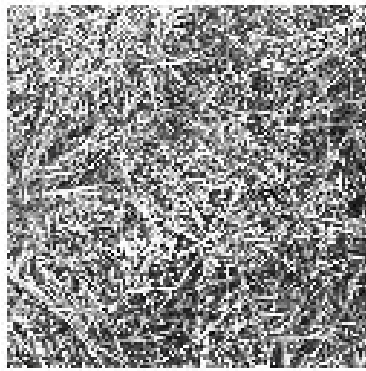
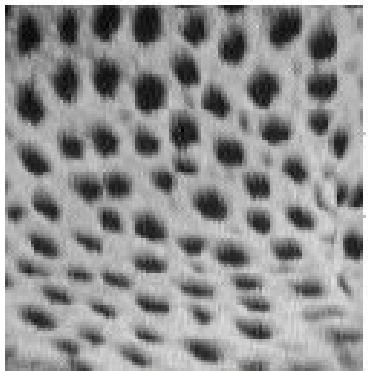
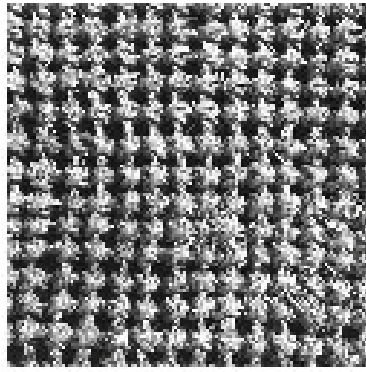
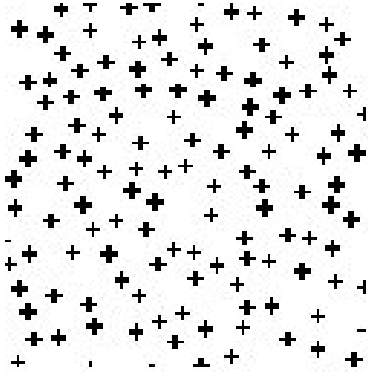
Variation Patterns:

1. **Expression** – geometric deformation
2. **illumination**--- Photometric deformation
3. **Transformation** –2D pose 3D
4. **Noise and Occlusion**



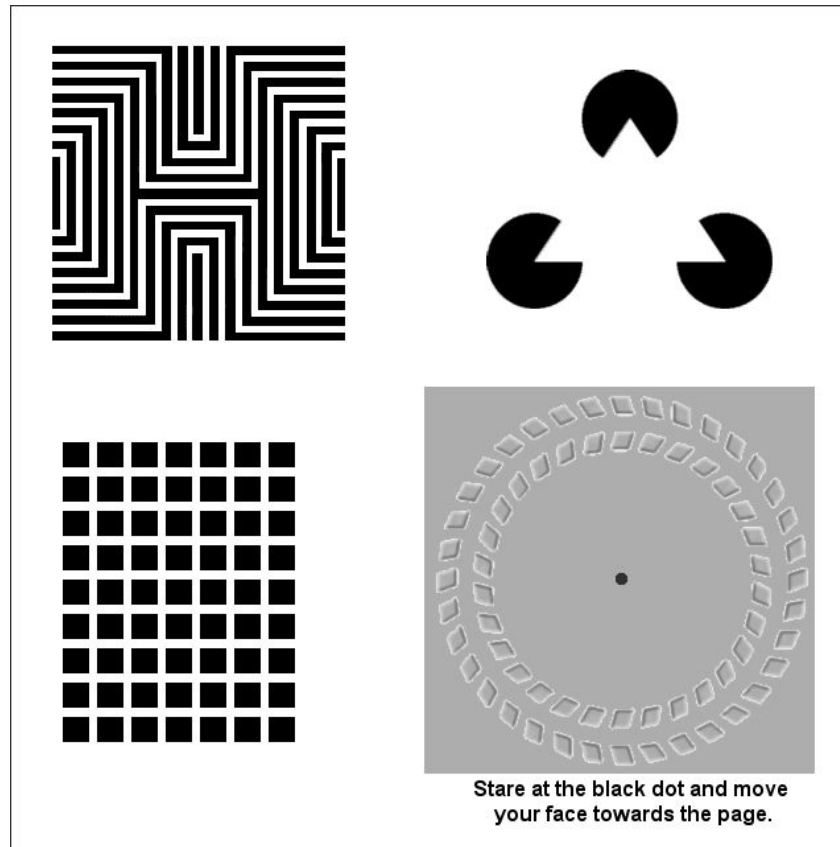
Examples of Patterns

A broad range of texture patterns are generated by stochastic processes.



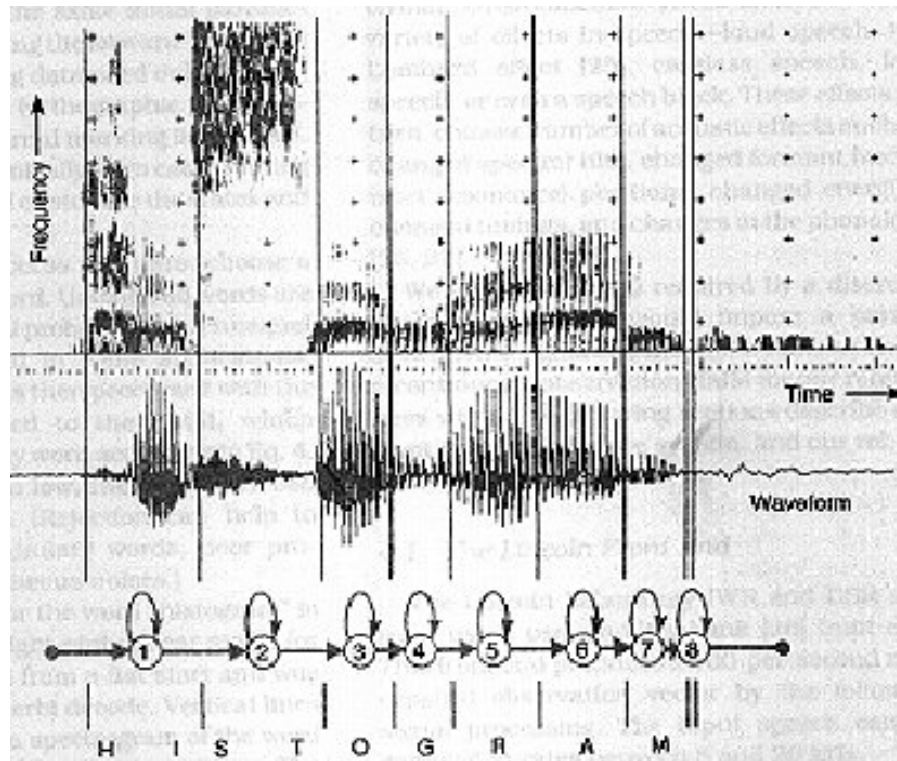
Examples of Patterns

How are these patterns represented in human mind?



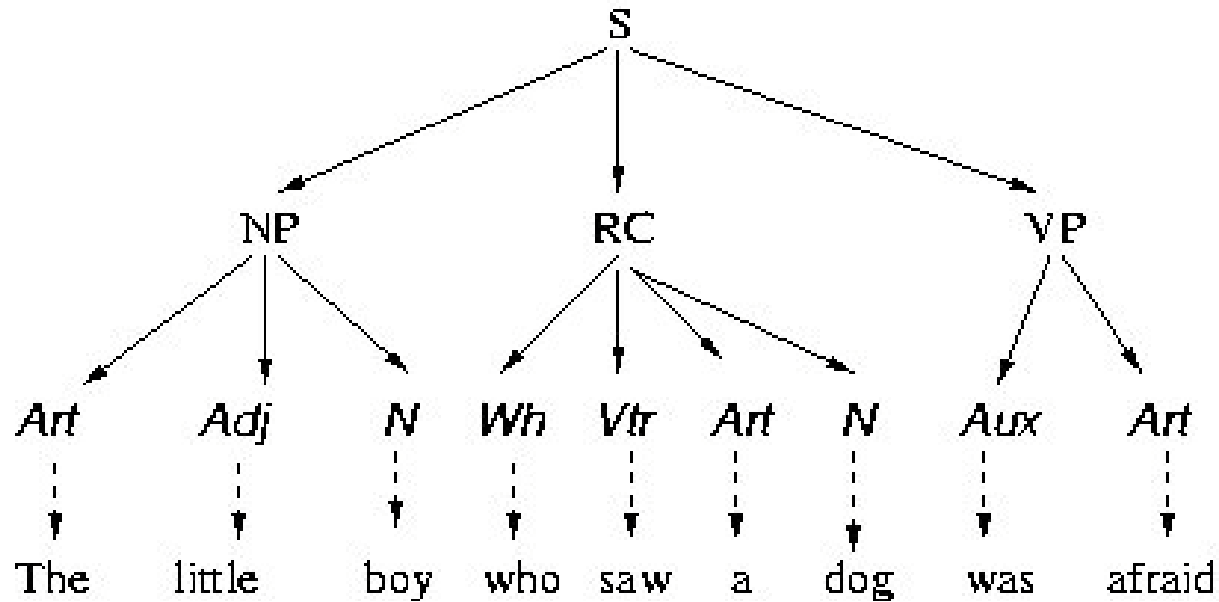
Examples of Patterns

Speech signals and Hidden Markov models



Examples of Patterns

Natural Language and stochastic grammar.

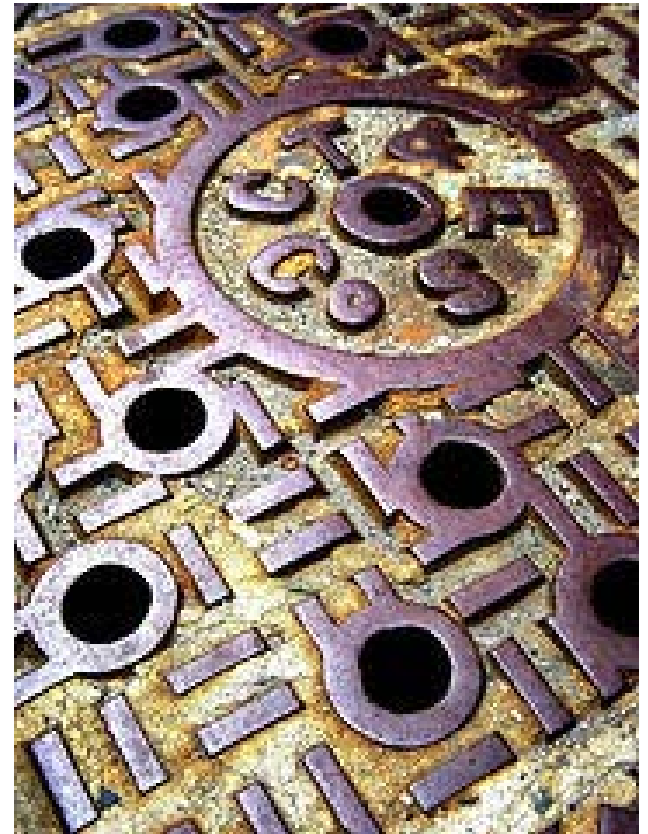


Examples of Patterns

Object Recognition

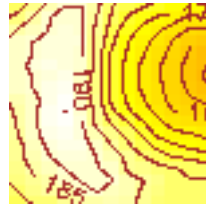


Patterns everywhere ?

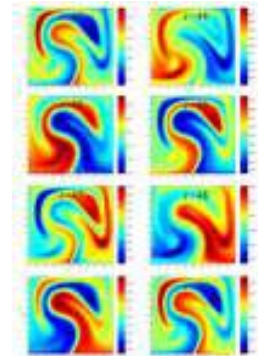
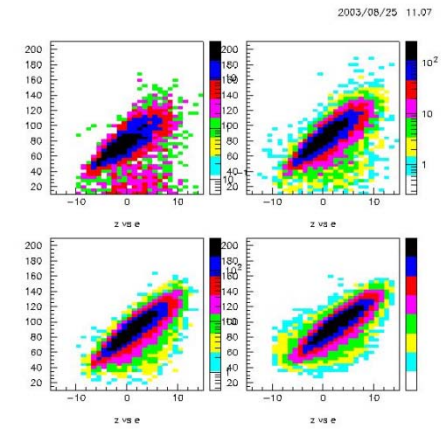


Examples of Patterns

Maps Recognition

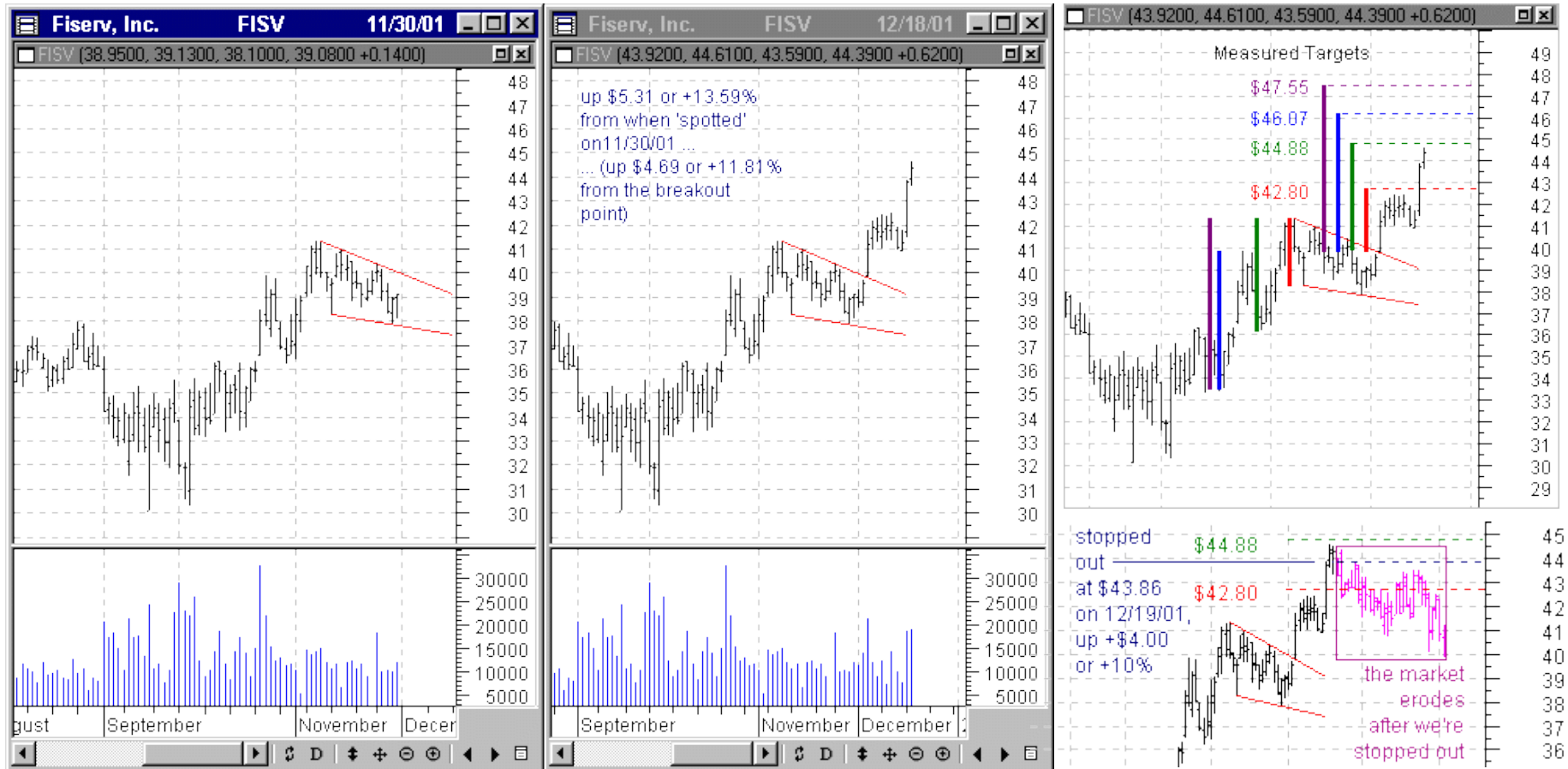


Patterns of Global Warming ?



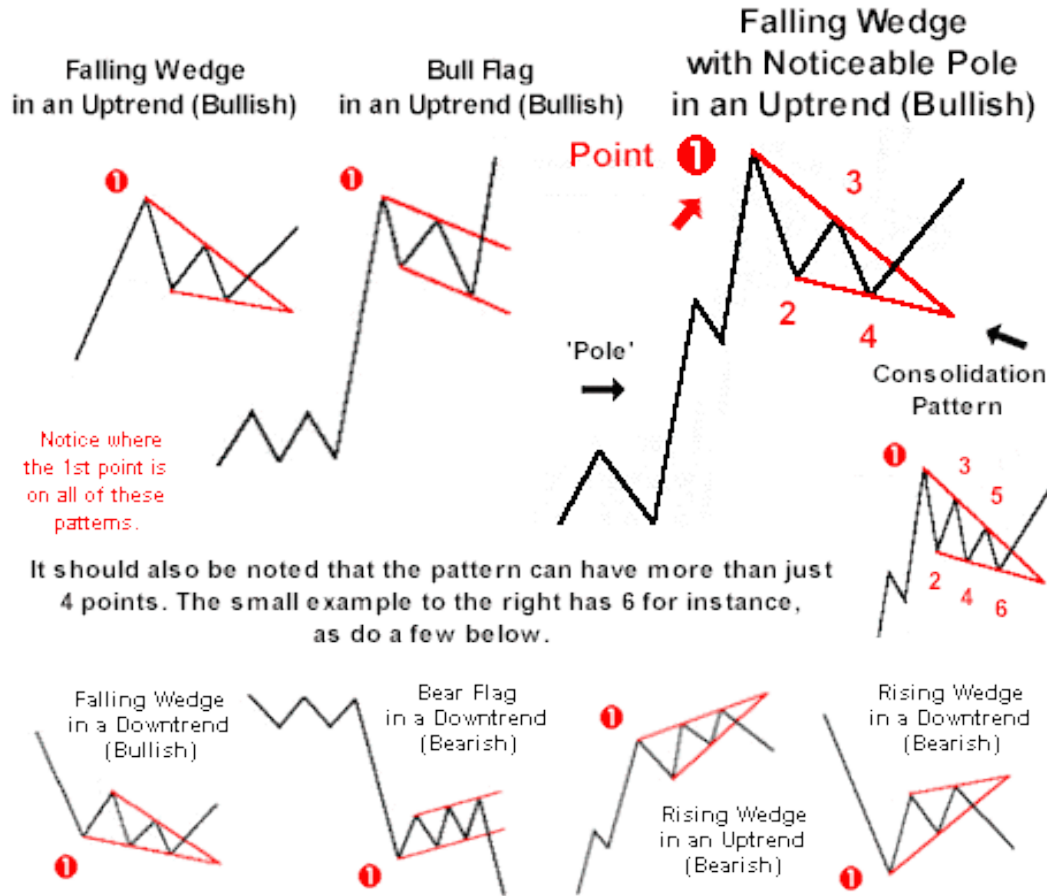
Examples of Patterns

Financial Series Pattern Recognition



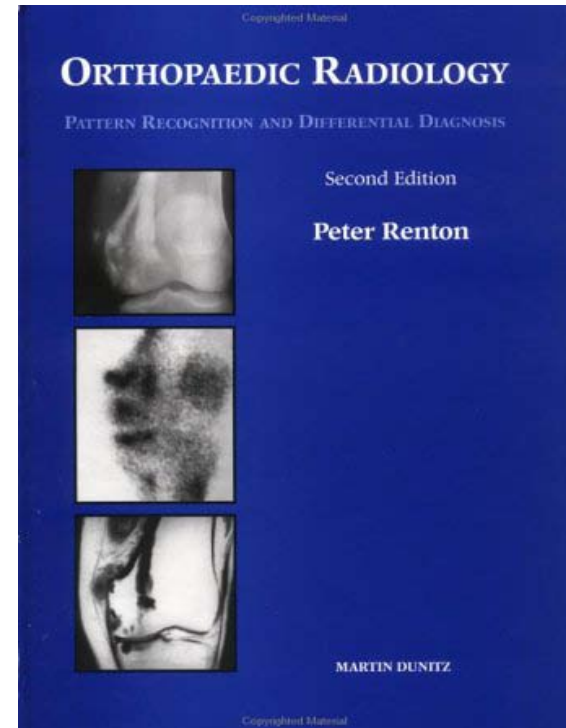
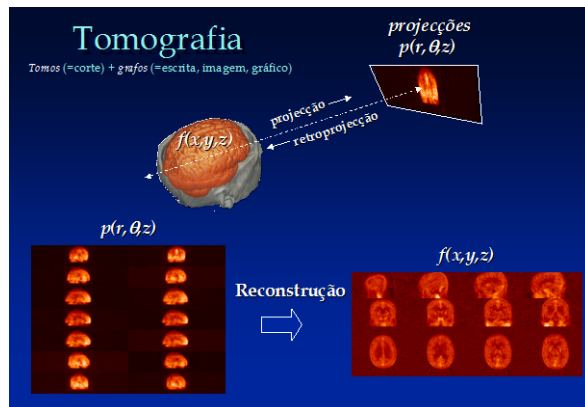
Examples of Patterns

How to Trade Chart Patterns ?



Examples of Patterns

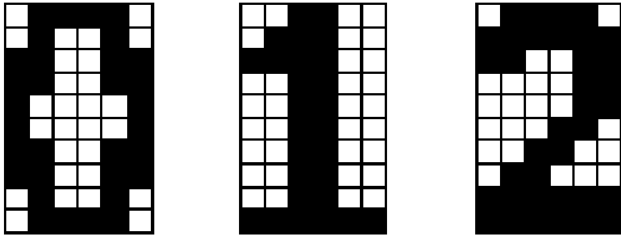
Pattern Recognition in Medical Diagnosis



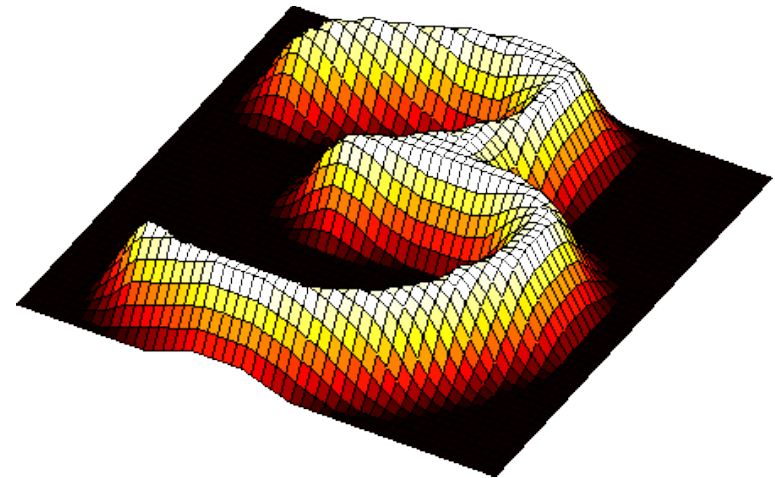
Examples of Patterns

Optical Character Recognition

**Padrões
a memorizar**



**Padrão a ser
reconhecido**



Examples of Patterns

Graphic Arts



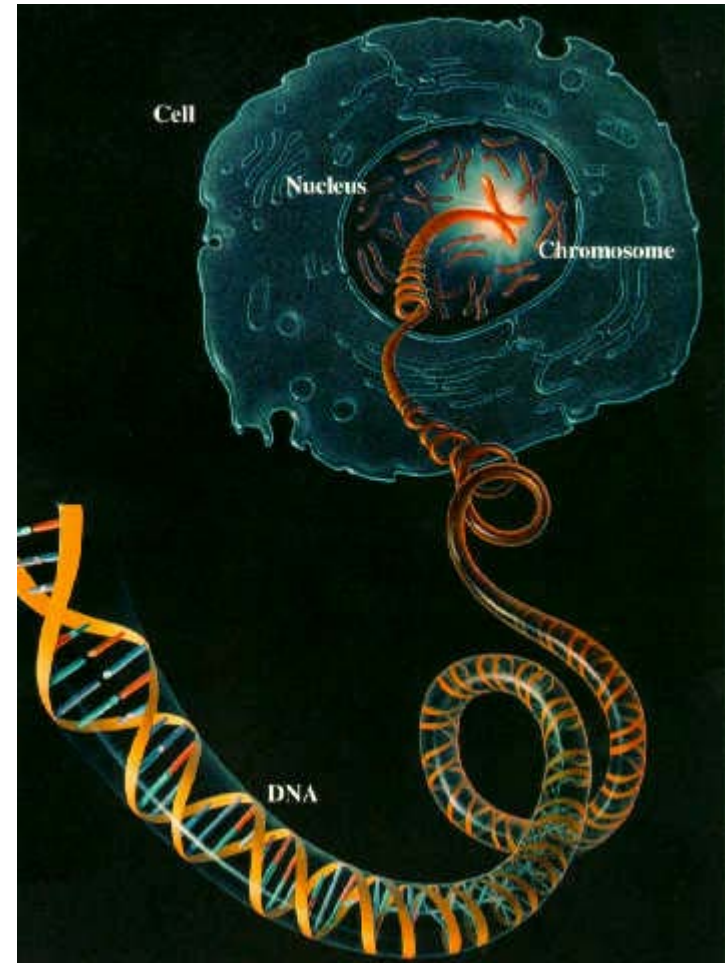
Escher, who else?

Examples of Patterns

Human Genome



Beautiful Patterns!



Approaches

- **Statistical PR:** based on underlying statistical model of patterns and pattern classes.
- **Neural networks:** classifier is represented as a network of cells modeling neurons of the human brain (connectionist approach).
- **Structural (or syntactic) PR:** pattern classes represented by means of formal structures as grammars, automata, strings, etc.

What is a pattern class?

- A **pattern class** (or category) is a set of patterns sharing common attributes.
- A collection of “similar” (not necessarily identical) objects.



- During **recognition** given objects are assigned to prescribed classes.

What is pattern recognition?

- Theory, Algorithms, Systems to put Patterns into Categories
- Relate Perceived Pattern to Previously Perceived Patterns
- Learn to distinguish patterns of interest from their background

Human Perception

- 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.



Examples of applications

Optical Character Recognition (OCR)

- Handwritten: sorting letters by postal code.
- Printed texts: reading machines for blind people, digitalization of text documents.

Biometrics

- Face recognition, verification, retrieval.
- Finger prints recognition.
- Speech recognition.

Diagnostic systems

- Medical diagnosis: X-Ray, EKG (ElectroCardioGraph) analysis.

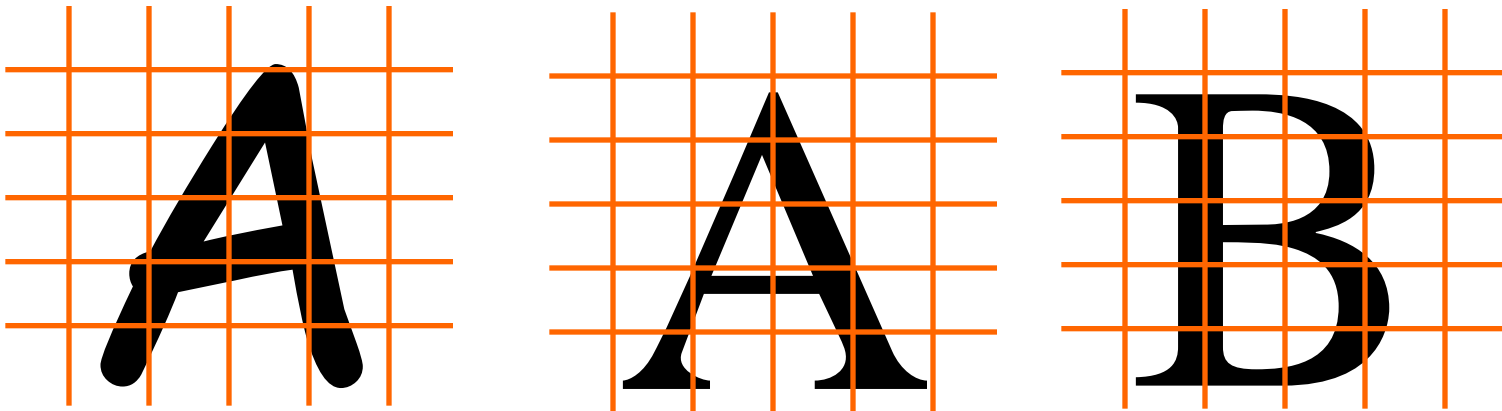
Military applications

- Automated Target Recognition (ATR).
- Image segmentation and analysis (recognition from aerial or satellite photographs).

THE STATISTICAL APPROACH

Examples of applications

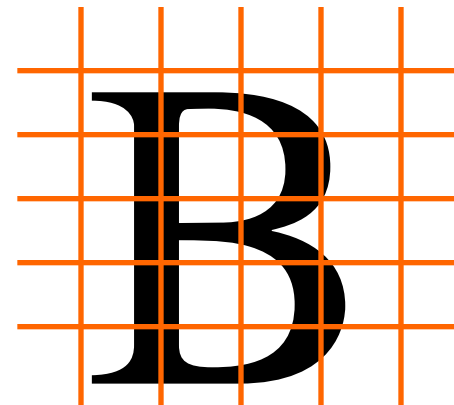
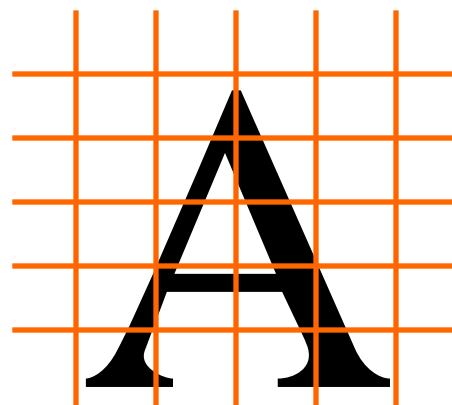
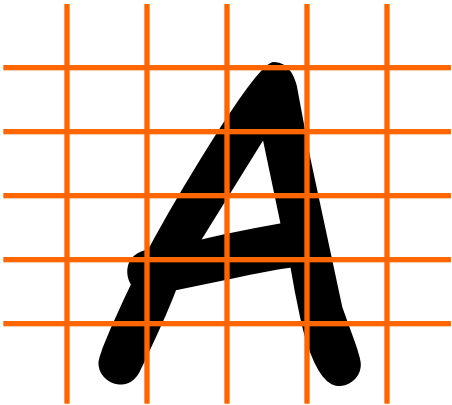
Grid by Grid Comparison



**Grid by Grid
Comparison**

Examples of applications

Grid by Grid Comparison



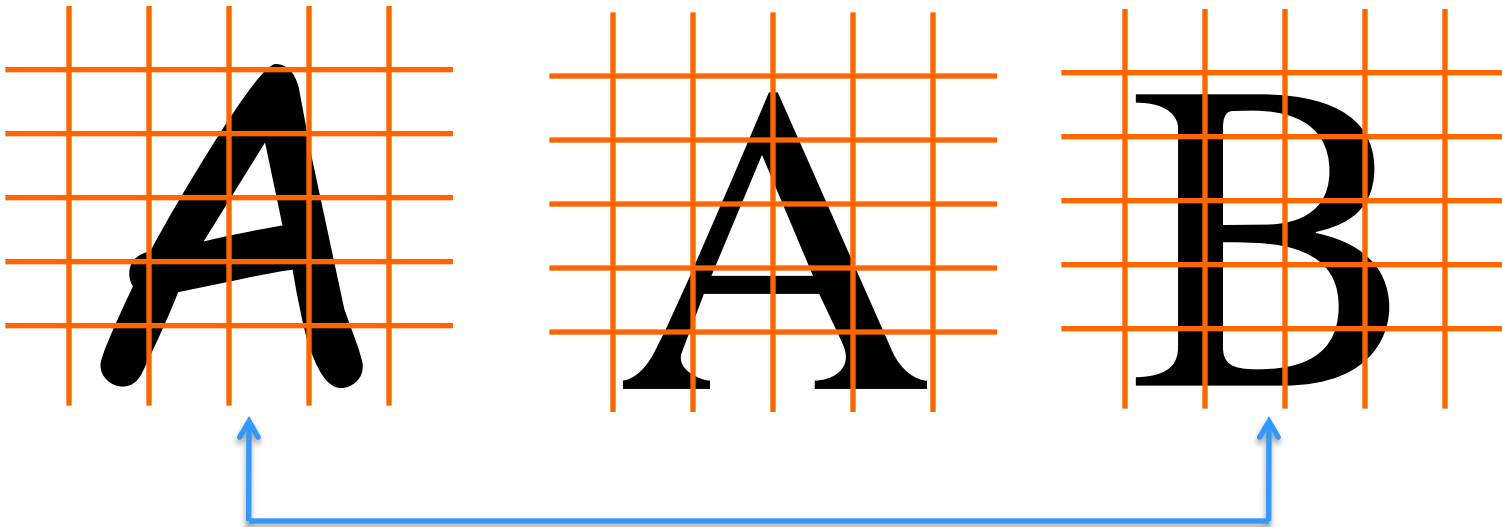
0	0	1	0
0	0	1	0
0	1	1	1
1	0	0	1
1	0	0	1

0	1	1	0
0	1	1	0
0	1	1	0
1	0	0	1
1	0	0	1

No of
Mismatch= 3

Examples of applications

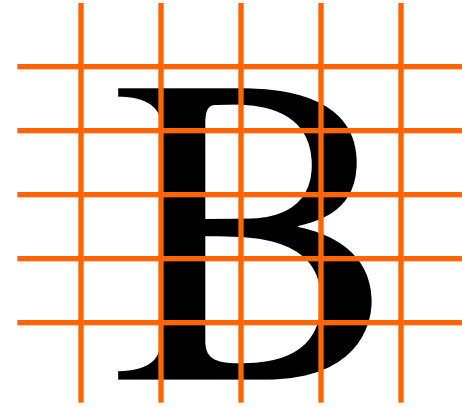
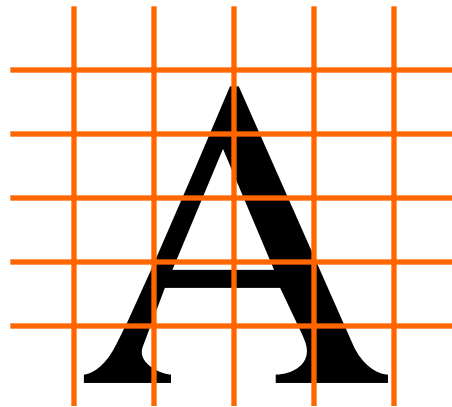
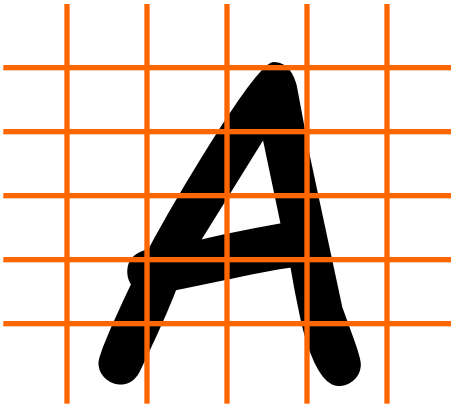
Grid by Grid Comparison



**Grid by Grid
Comparison**

Examples of applications

Grid by Grid Comparison



0	0	1	0
0	0	1	0
0	1	1	1
1	0	0	1
1	0	0	1

No of
Mismatch= 9

1	1	1	0
0	1	0	1
0	1	1	1
0	1	0	1
1	1	1	0

* / - + 1 @ #

Problem with Grid by Grid Comparison

- *Time to recognize a pattern* - Proportional to the number of stored patterns (Too costly with the increase of number of patterns stored



Human and Machine Perception

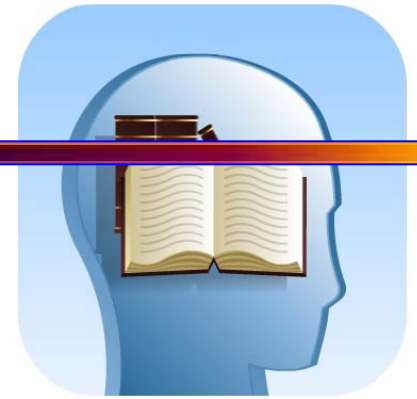
- We are often influenced by the knowledge of how patterns are modeled and recognized in nature when we develop pattern recognition algorithms.
- Research on machine perception also helps us gain deeper understanding and appreciation for pattern recognition systems in nature.
- Yet, we also apply many techniques that are purely numerical and do not have any correspondence in natural systems.

Pattern Recognition



- Two Phase : ***Learning*** and ***Detection***.
- Time to learn is higher.
 - *Driving a car*
- Difficult to learn but once learnt it becomes **natural**.
- Can use AI learning methodologies such as:
 - Neural Network.
 - Machine Learning.

Learning



- How can machine learn the rule from data?
 - **Supervised learning:** a teacher provides a category label or cost for each pattern in the training set.
 - Here samples from information classes (training data) are used for learning and then classifying unknown data points/patterns.
 - **Unsupervised learning:** the system forms clusters or natural groupings of the input patterns.
 - **Semi-Supervised:** A small training samples from information classes are used for initial learning. The model is further built using unlabeled samples for classifying unknown patterns

Types of Prediction Problems (1/2)

Classification

- The PR problem of assigning an object to a class
- The output of the PR system is an integer label
 - e.g. classifying a product as “good” or “bad” in a quality control test

Regression

- A generalization of a classification task
- The output of the PR system is a real-valued number
 - e.g. predicting the share value of a firm based on past performance and stock market indicators

Types of Prediction Problems (2/2)

Clustering

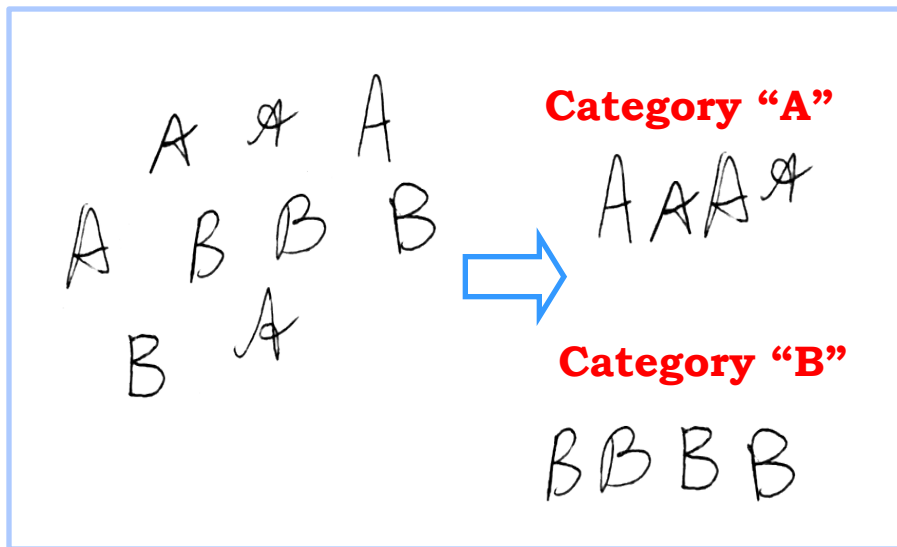
- The problem of organizing objects into meaningful groups
- The system returns a (sometimes hierarchical) grouping of objects
 - e.g. organizing life forms into a taxonomy of species

Description

- The problem of representing an object in terms of a series of primitives
- The PR system produces a structural or linguistic description
 - e.g. labeling an ECG signal in terms of P, QRS and T complexes

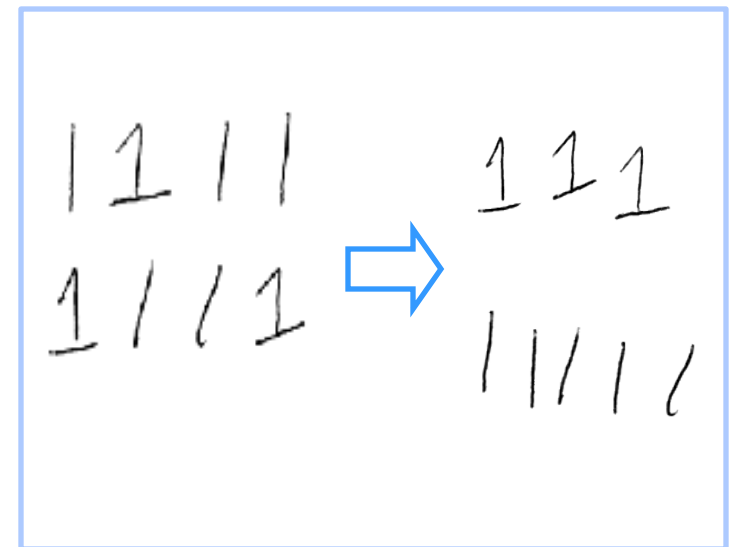
Classification vs. Clustering

- *Classification* (known categories)
- *Clustering* (creation of new categories)



Classification

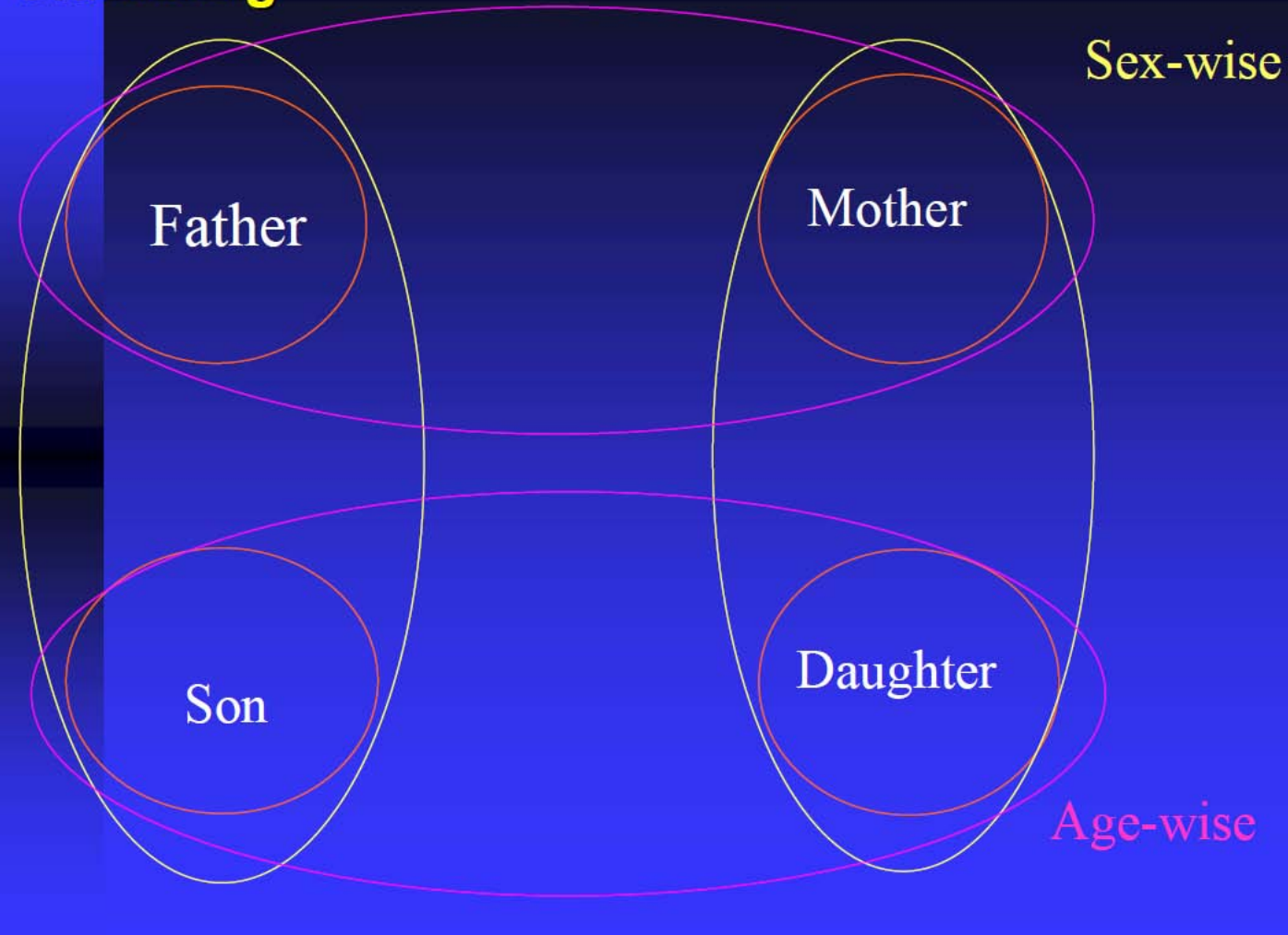
(Supervised Classification)



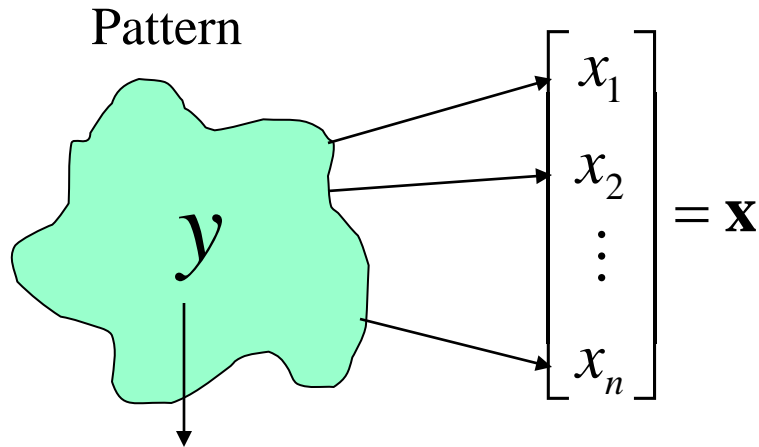
Clustering

(Unsupervised Classification)

Clustering



Basic concepts (Classification)



Feature vector $\mathbf{x} \in X$

- A vector of observations (measurements).
- \mathbf{x} is a point in feature space X .

Hidden state $y \in Y$

- Cannot be directly measured.
- Patterns with equal hidden state belong to the same class.

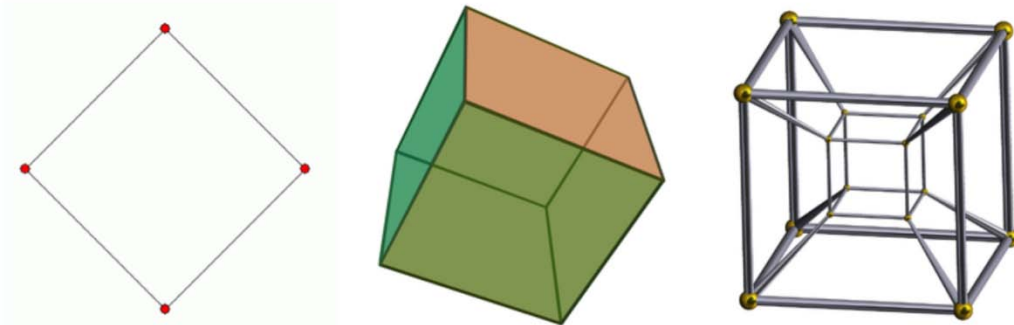
Task

- To design a classifier (decision rule) $d : X \rightarrow Y$ which decides about a hidden state based on an onbservation.

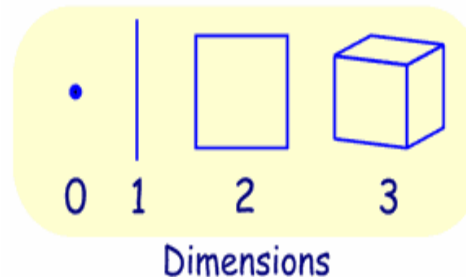
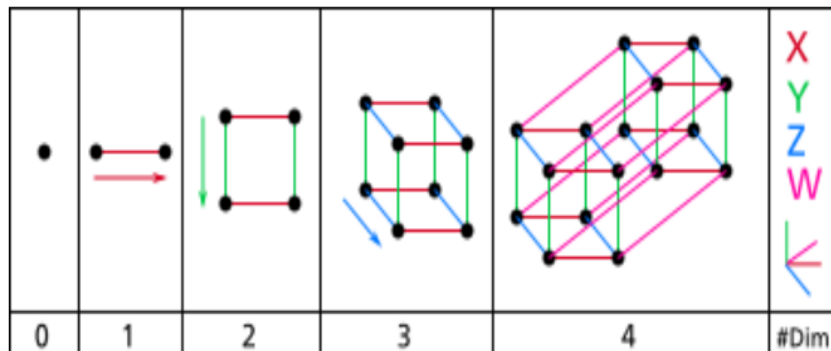
Dimension

The dimension of a space or object is informally defined as the minimum number of coordinates needed to specify any point within it, e.g.,

- A line has a dimension of one because only one coordinate is needed to specify a point on it (for example, the point at 5 on a number line).
- A surface such as a plane or the surface of a cylinder or sphere has a dimension of two because two coordinates are needed to specify a point on it (for example, to locate a point on the surface of a sphere you need both its latitude and its longitude).
- The inside of a cube, a cylinder or a sphere is three-dimensional because three coordinates are needed to locate a point within these spaces.



From left to right, the [square](#), the [cube](#), and the [tesseract](#). The square is bounded by 1-dimensional lines, the cube by 2-dimensional areas, and the tesseract by 3-dimensional volumes.



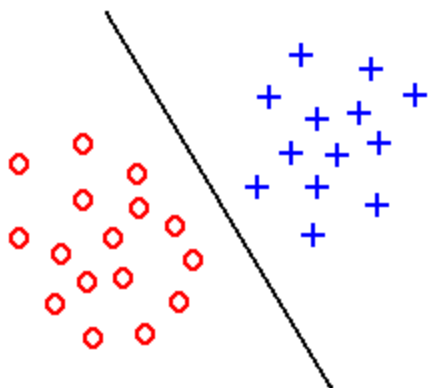
Point, Line, Plane and Solid

- A [point](#) has no dimensions, only location
- A [line](#) is one-dimensional
- A [plane](#) is two dimensional
- A [solid](#) is three-dimensional

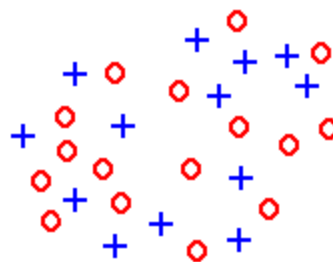
Feature extraction

Task: to extract features which are good for classification.

- Good features:
- Objects from the same class have similar feature values.
 - Objects from different classes have different values.



“Good” features



“Bad” features

Dimensionality Reduction

Dimensionality Reduction

- A limited yet **salient feature set** simplifies both **pattern representation** and **classifier design**.
- **Pattern representation** is easy for **2D** and **3D** features.
- How to make pattern with **high dimensional** features **viewable**?

Dimensionality Reduction How?

- Feature Extraction
 - Create new features based on the original feature set
 - Transforms are usually involved
- Feature Selection
 - Select the best subset from a given feature set.

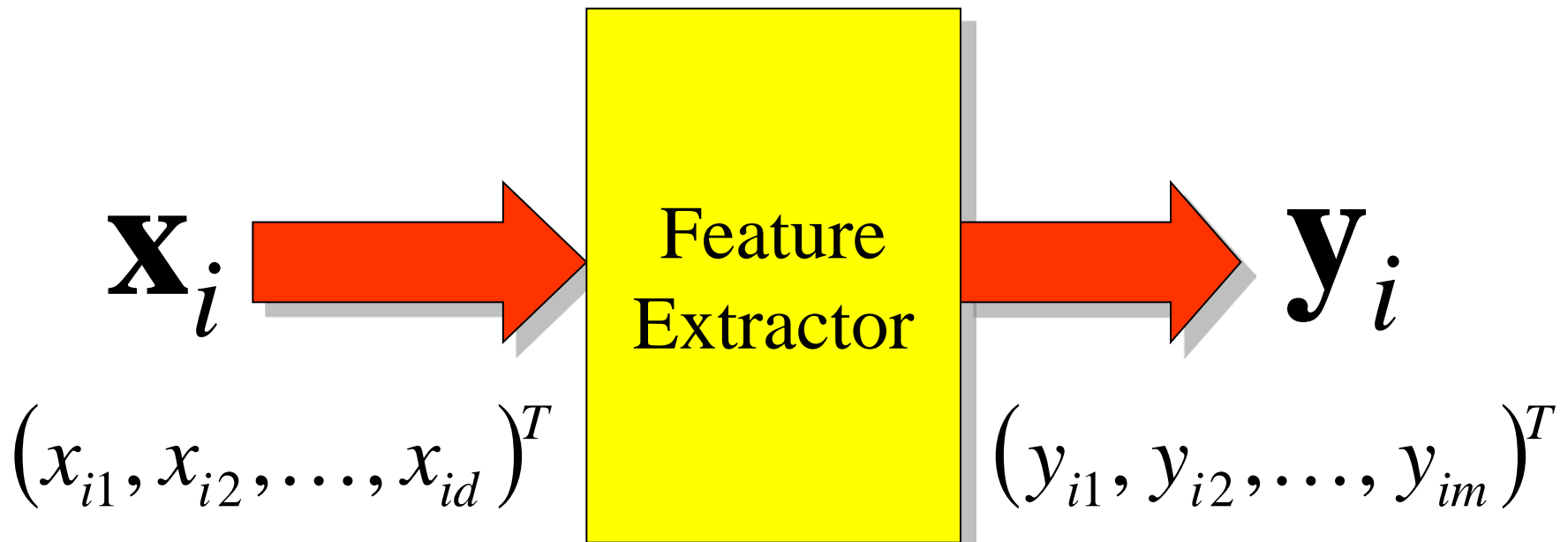
Main Issues in Dimensionality Reduction

- The choice of a **criterion function**
 - Commonly used criterion: **classification error**
- The determination of the **appropriate dimensionality**
 - Correlated with the **intrinsic dimensionality** of data

Dimensionality Reduction

Feature Extraction

Feature Extractor



$m \leq d$, usually

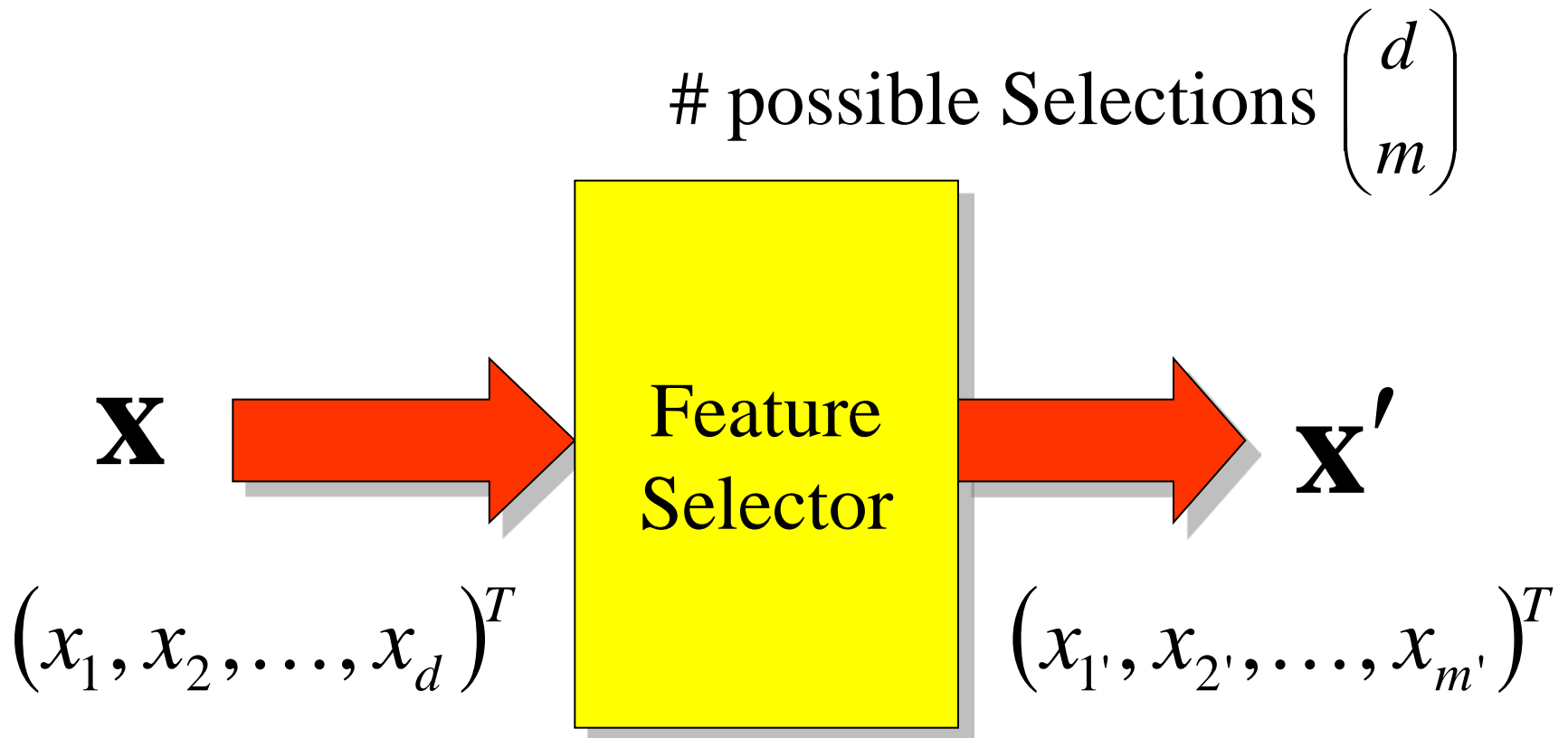
Some Important Methods

- Principal Component Analysis (PCA)
 - or Karhunen-Loeve Expansion
 - Project Pursuit
 - Independent Component Analysis (ICA)
 - Factor Analysis
 - Discriminate Analysis
- } Linear Approaches
- Kernel PCA
 - Multidimensional Scaling (MDS)
- } Nonlinear Approaches
- Feed-Forward Neural Networks
 - Self-Organizing Map
- } Neural Networks

Dimensionality Reduction

Feature Selection

Feature Selector



$m \leq d$, usually

The problem

- Given a set of d features, select a subset of size m that leads to **the smallest classification error**.
- **No nonexhaustive** sequential feature selection procedure can be guaranteed to produce the optimal subset.

Optimal Methods

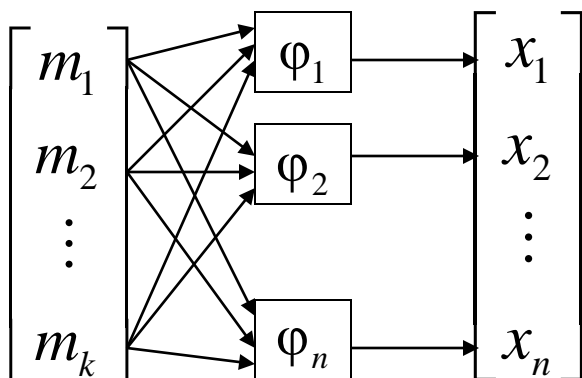
- Exhaustive Search
 - Evaluate all possible subsets
- Branch-and-Bound
 - The **monotonicity property** of the criterion function has to be held.

Suboptimal Methods

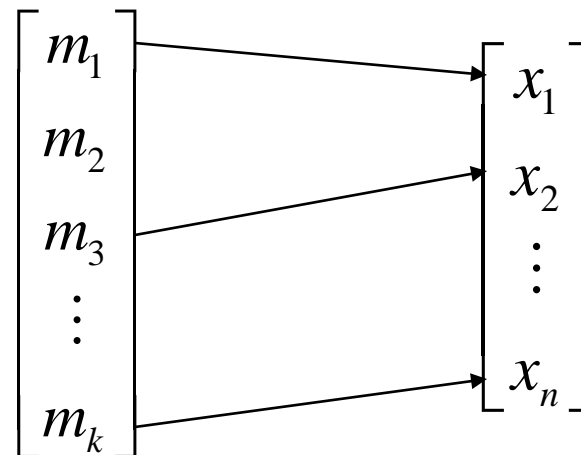
- Best **Individual** Features
- Sequential **Forward** Selection (SFS)
- Sequential **Backward** Selection (SBS)
- “**Plus l -take away r** ” Selection
- Sequential **Forward Floating** Search and Sequential **Backward Floating** Search

Feature extraction methods

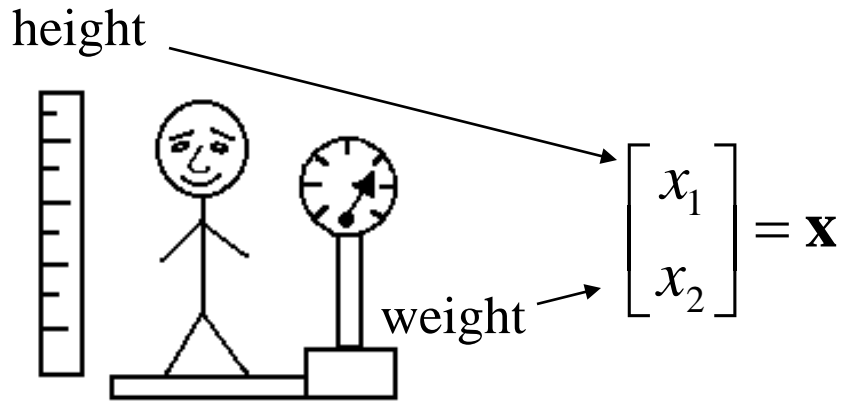
Feature extraction



Feature selection



Example: 1



Task: jockey-hoopster recognition.

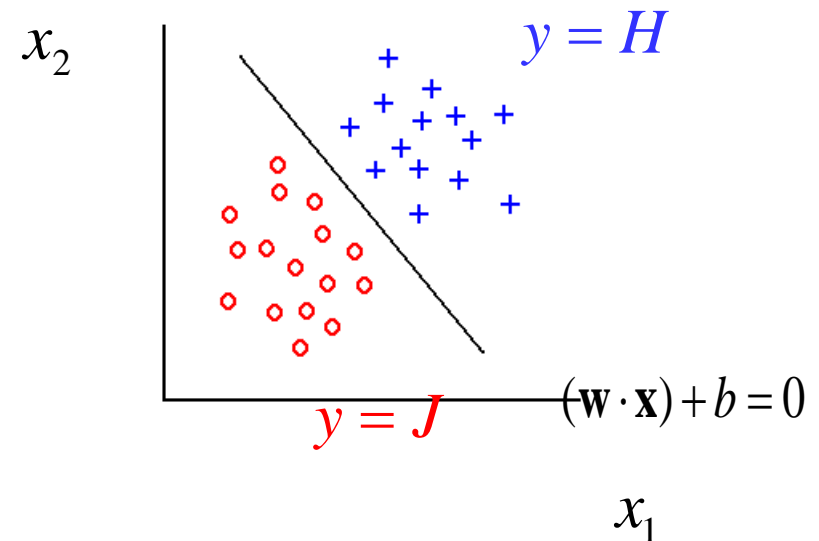
The set of hidden state is $Y = \{H, J\}$

The feature space is $X = \mathfrak{R}^2$

Training examples $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$

Linear classifier:

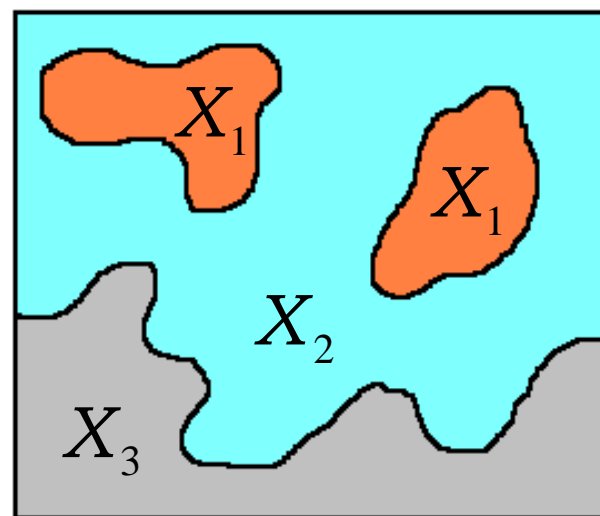
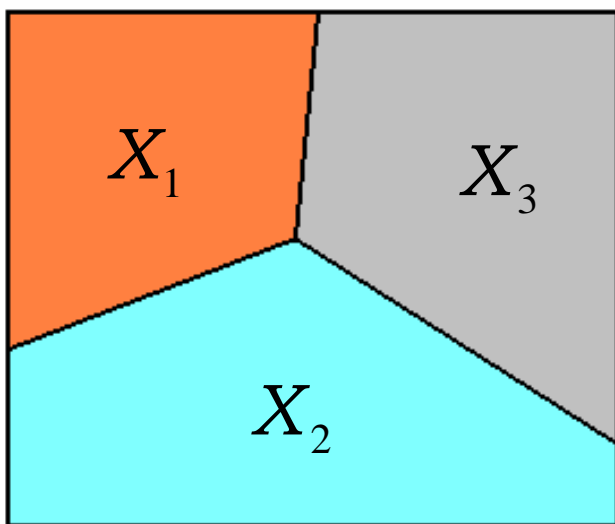
$$d(\mathbf{x}) = \begin{cases} H & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b \geq 0 \\ J & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b < 0 \end{cases}$$



Classifier

A classifier partitions feature space X into **class-labeled regions** such that

$$X = X_1 \cup X_2 \cup \dots \cup X_{|Y|} \quad \text{and} \quad X_1 \cap X_2 \cap \dots \cap X_{|Y|} = \{0\}$$



The classification consists of determining to which region a feature vector \mathbf{x} belongs to.

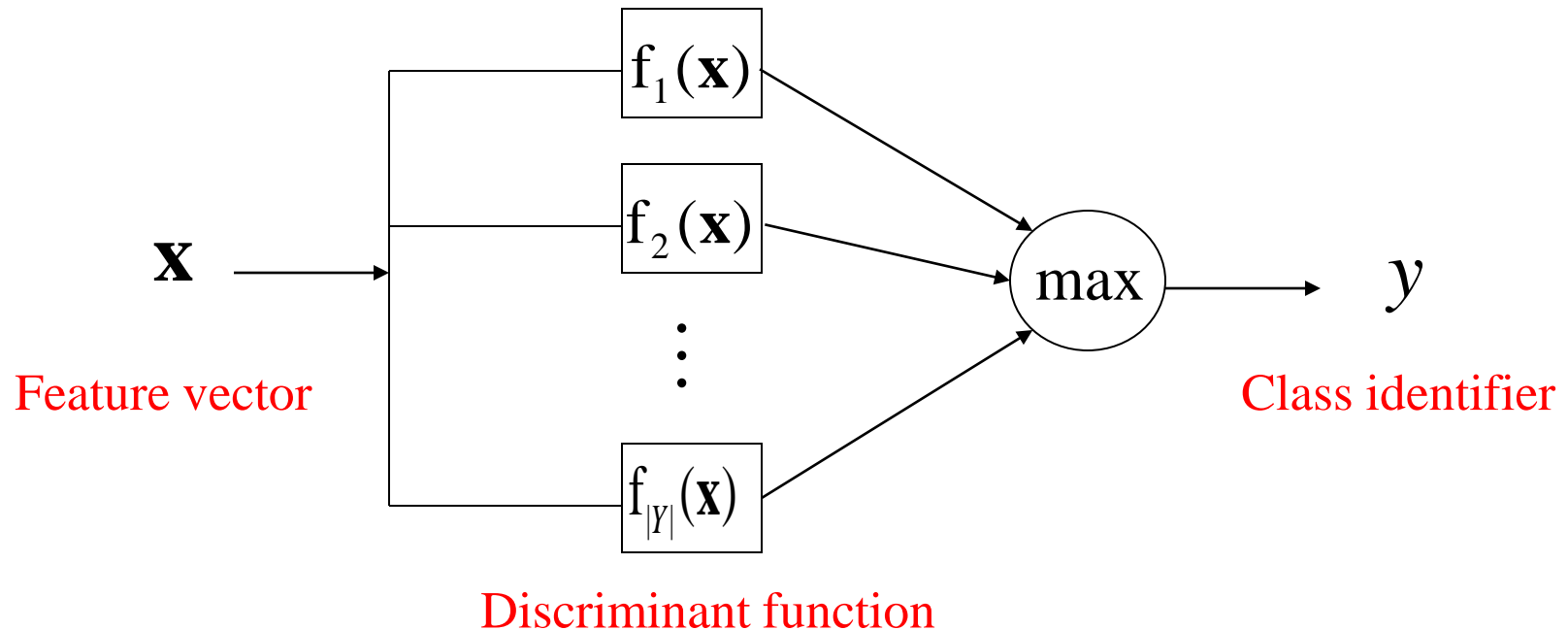
Borders between **decision boundaries** are called decision regions.

Representation of classifier

A classifier is typically represented as a set of discriminant functions

$$f_i(\mathbf{x}) : X \rightarrow \mathcal{R}, i = 1, \dots, |Y|$$

The classifier assigns a feature vector \mathbf{x} to the i -th class if $f_i(\mathbf{x}) > f_j(\mathbf{x}) \quad \forall j \neq i$



Pattern Discrimination Revisited

Decision Regions and Functions

The feature space is

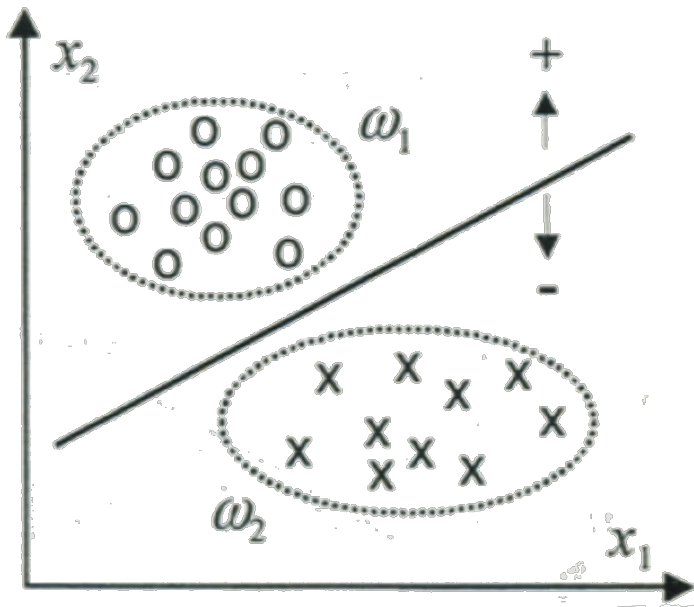
$$\mathbf{x} = [x_1, x_2] \in \mathcal{R}^2$$

Linear decision function :

$$d(\mathbf{x}) = w_1x_1 + w_2x_2 + w_0 = 0$$

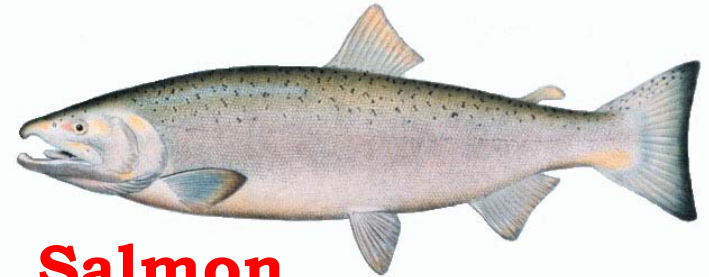
Linear classifier:

$$d(\mathbf{x}) = \begin{cases} o & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b \geq 0 \\ x & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b < 0 \end{cases}$$



Case Study

- Fish Classification:
 - Sea Bass / Salmon.



Salmon

- **Problem**: Sorting incoming fish on a conveyor belt according to species.

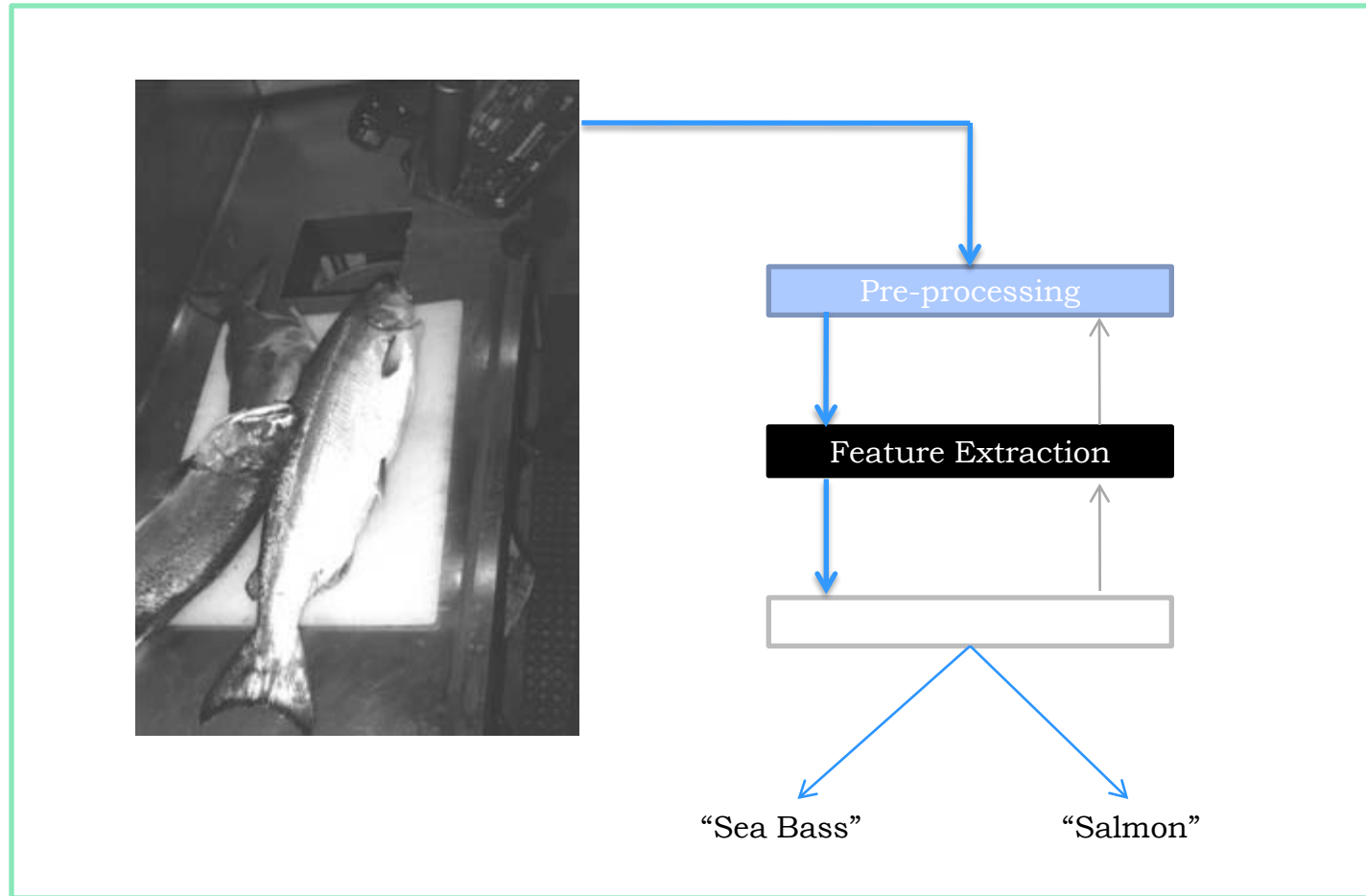


Sea-bass

Case Study (Cont.)

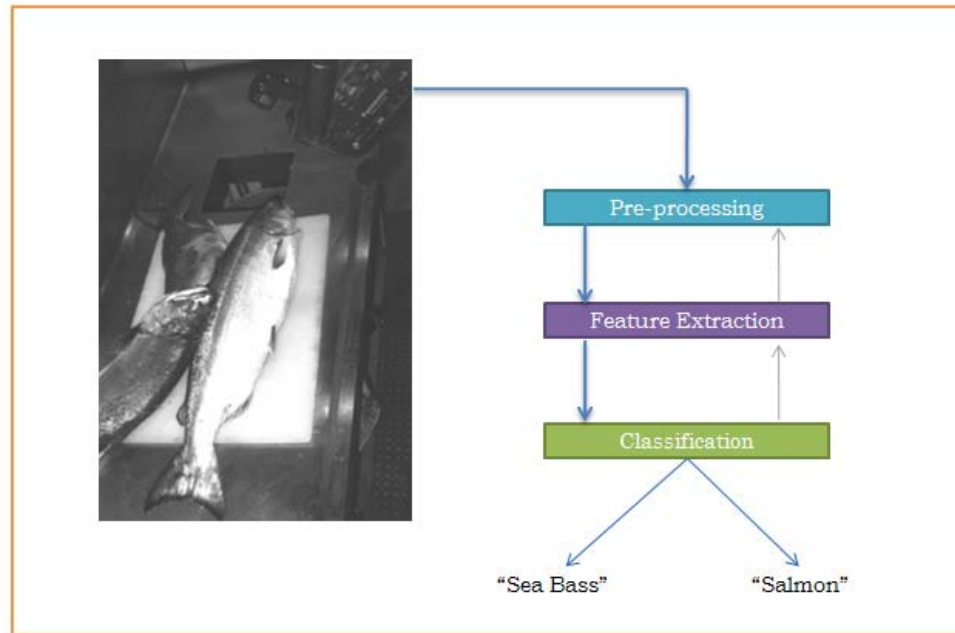
- What can cause problems during sensing?
 - Lighting conditions.
 - Position of fish on the conveyor belt.
 - Camera noise.
 - etc...
- What are the steps in the process?
 1. *Capture image.*
 2. *Isolate fish*
 3. *Take measurements*
 4. *Make decision*

Case Study (Cont.)



Case Study (Cont.)

- Pre-Processing:
 - Image enhancement
 - Separating touching or occluding fish.
 - Finding the boundary of the fish.

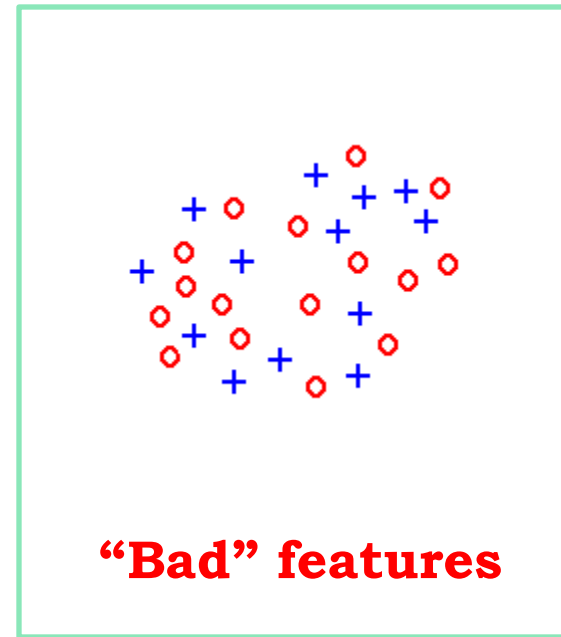
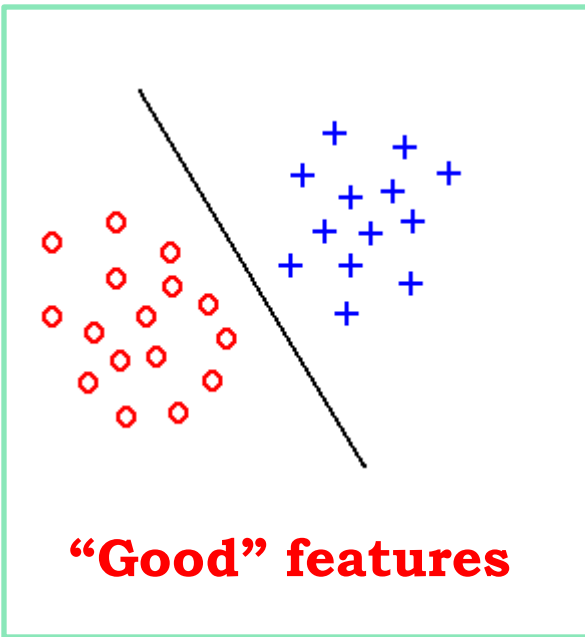


Case Study (Cont.)

How to separate sea bass from salmon?

- Possible features to be used:
 - Length
 - Lightness
 - Width
 - Number and shape of fins
 - Position of the mouth
 - Etc ...
- Assume a fisherman told us that a “sea bass” is generally longer than a “salmon”.
- Even though “sea bass” is longer than “salmon” on the average, there are many examples of fish where this observation does not hold.

Feature Selection: Good/Bad



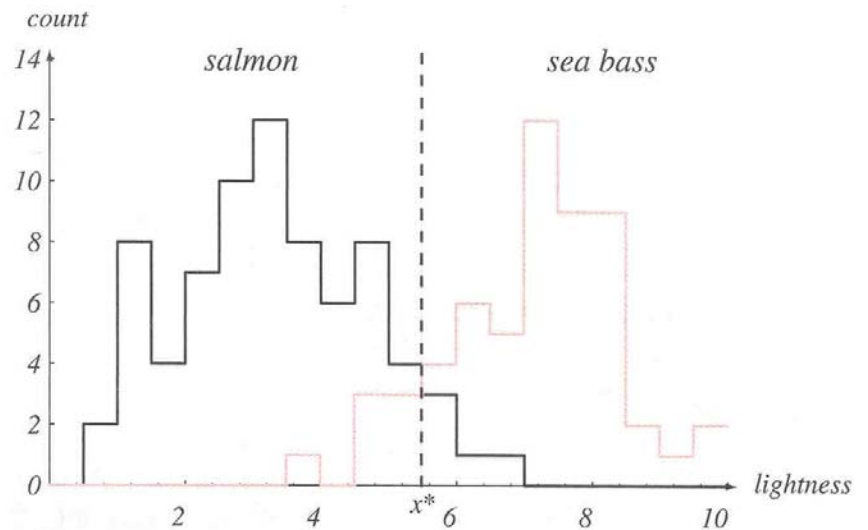
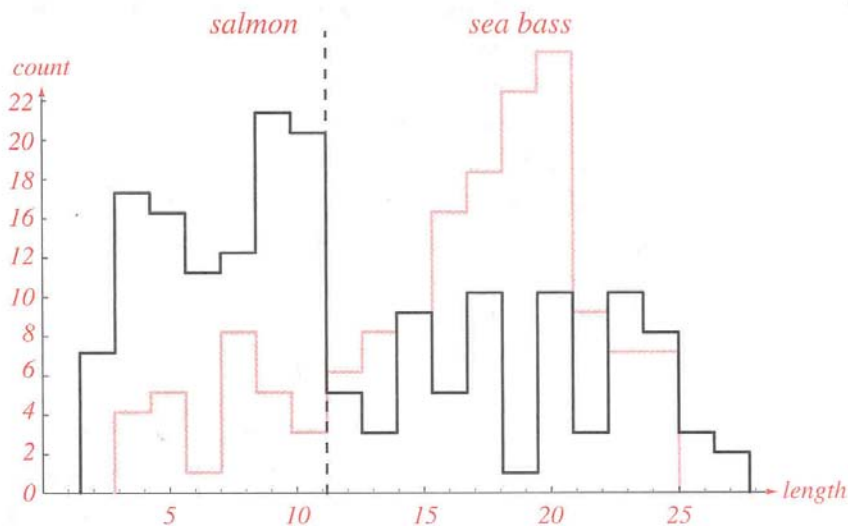
Case Study (Cont.)

How to separate sea bass from salmon?

- To improve recognition, we might have to use more than one feature at a time.
 - Single features might not yield the best performance.
 - Combinations of features might yield better performance.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \begin{array}{l} x_1 : \textit{lightness} \\ x_2 : \textit{width} \end{array}$$

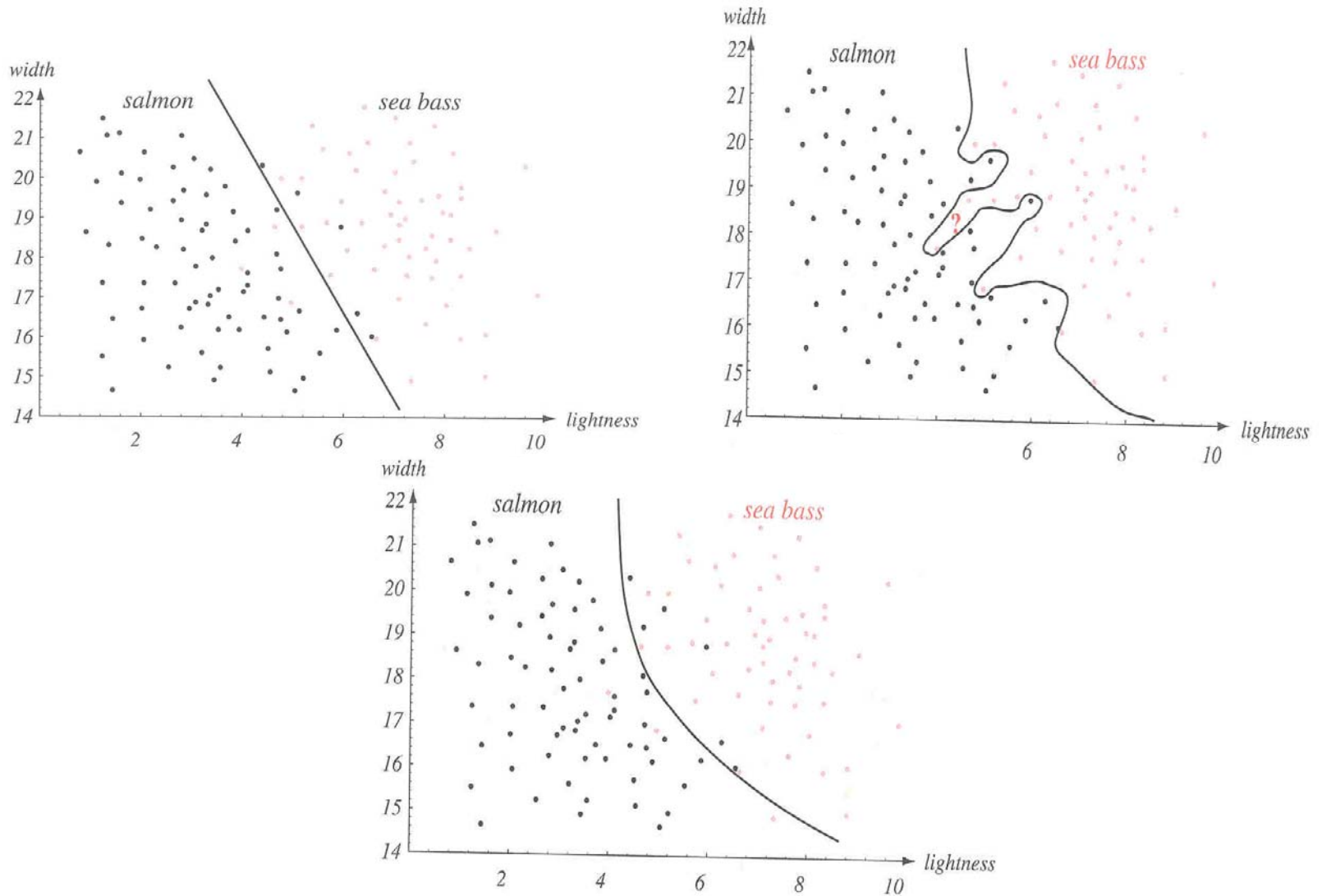
Features and Distributions



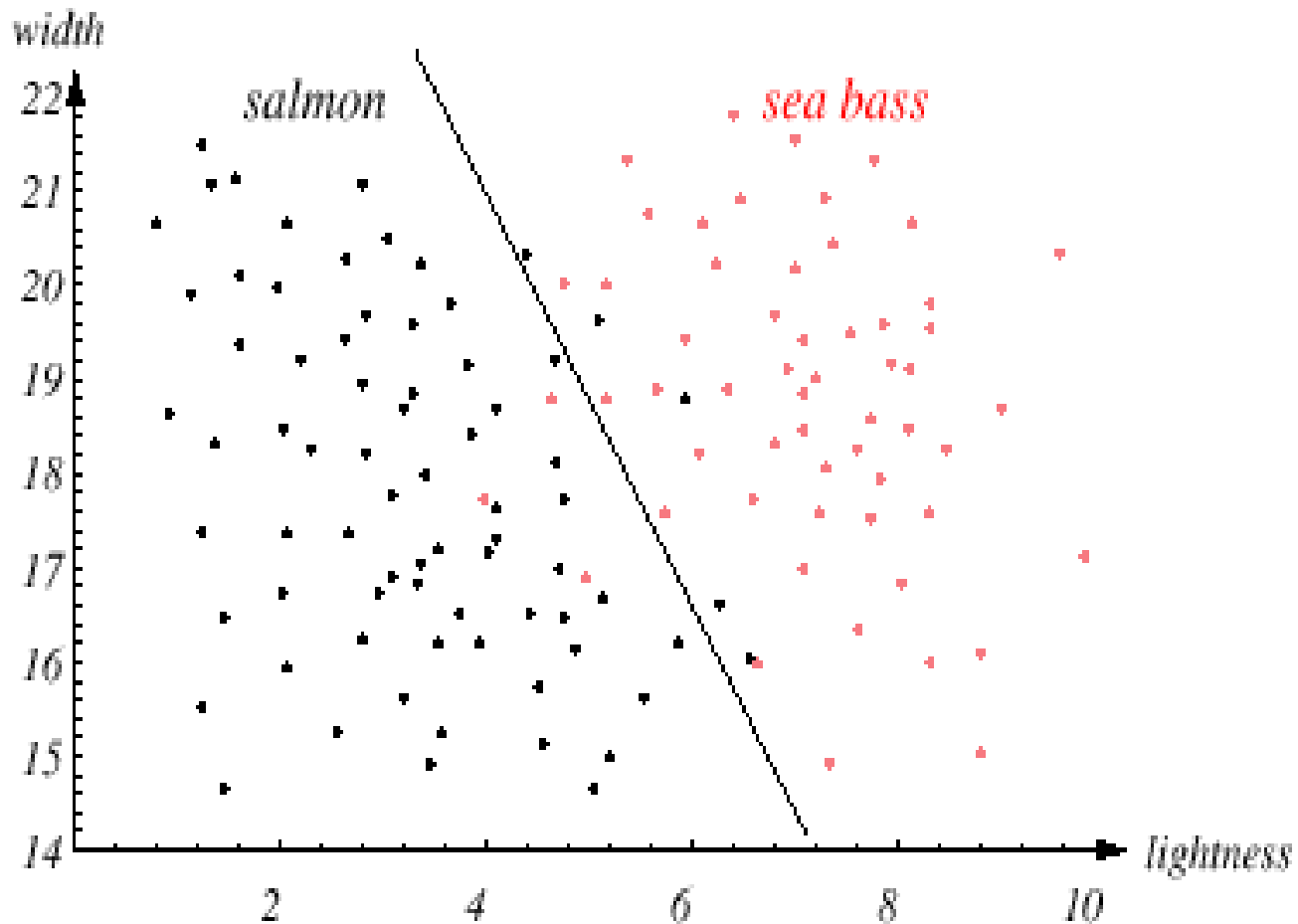
The **length/width** is a poor feature alone!

Select the **lightness** as a possible feature.

Decision/classification Boundaries



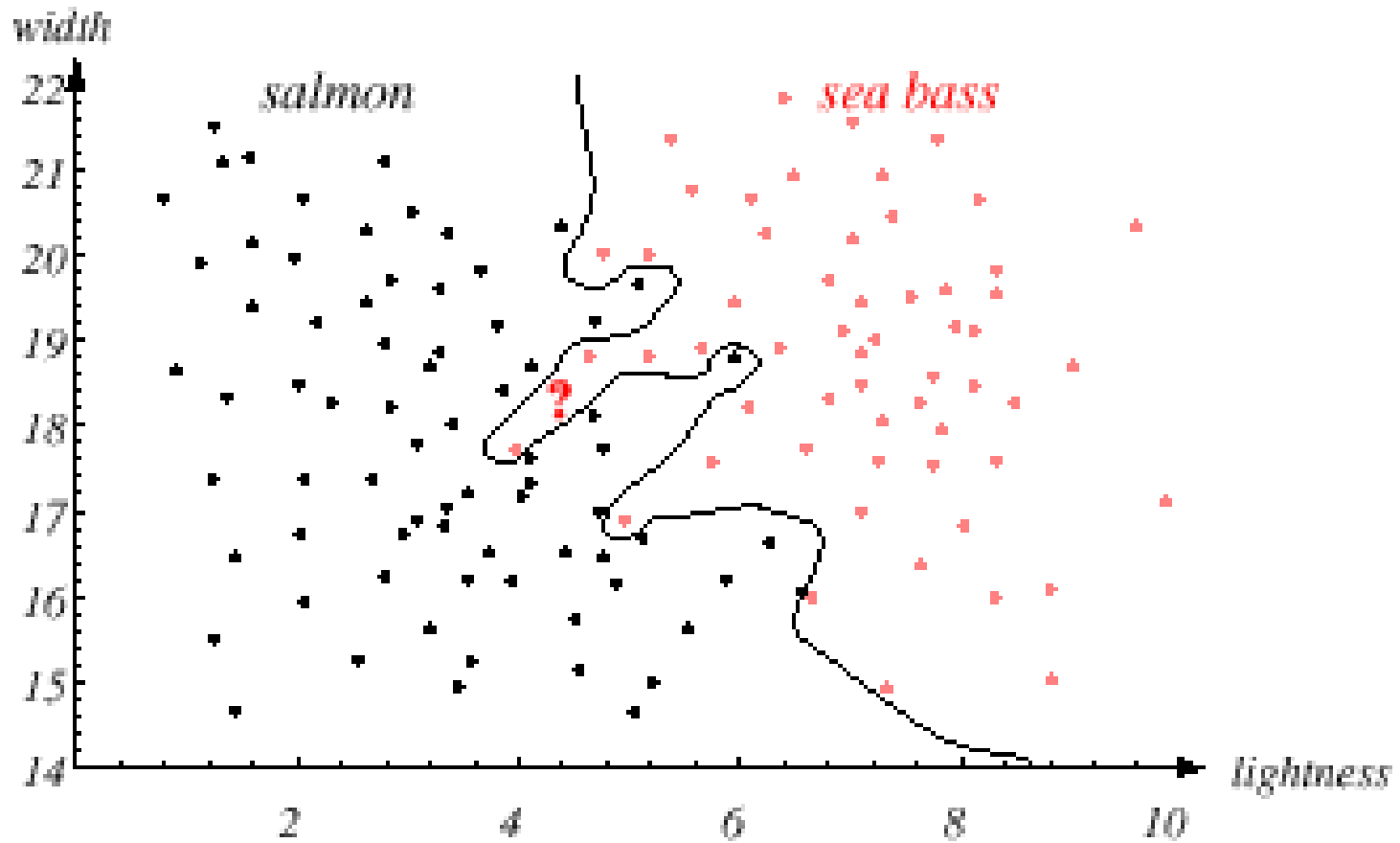
Decision/classification Boundaries



Decision/classification Boundaries

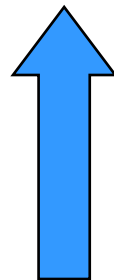
- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such “noisy features”
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure (next slide):

Decision/classification Boundaries



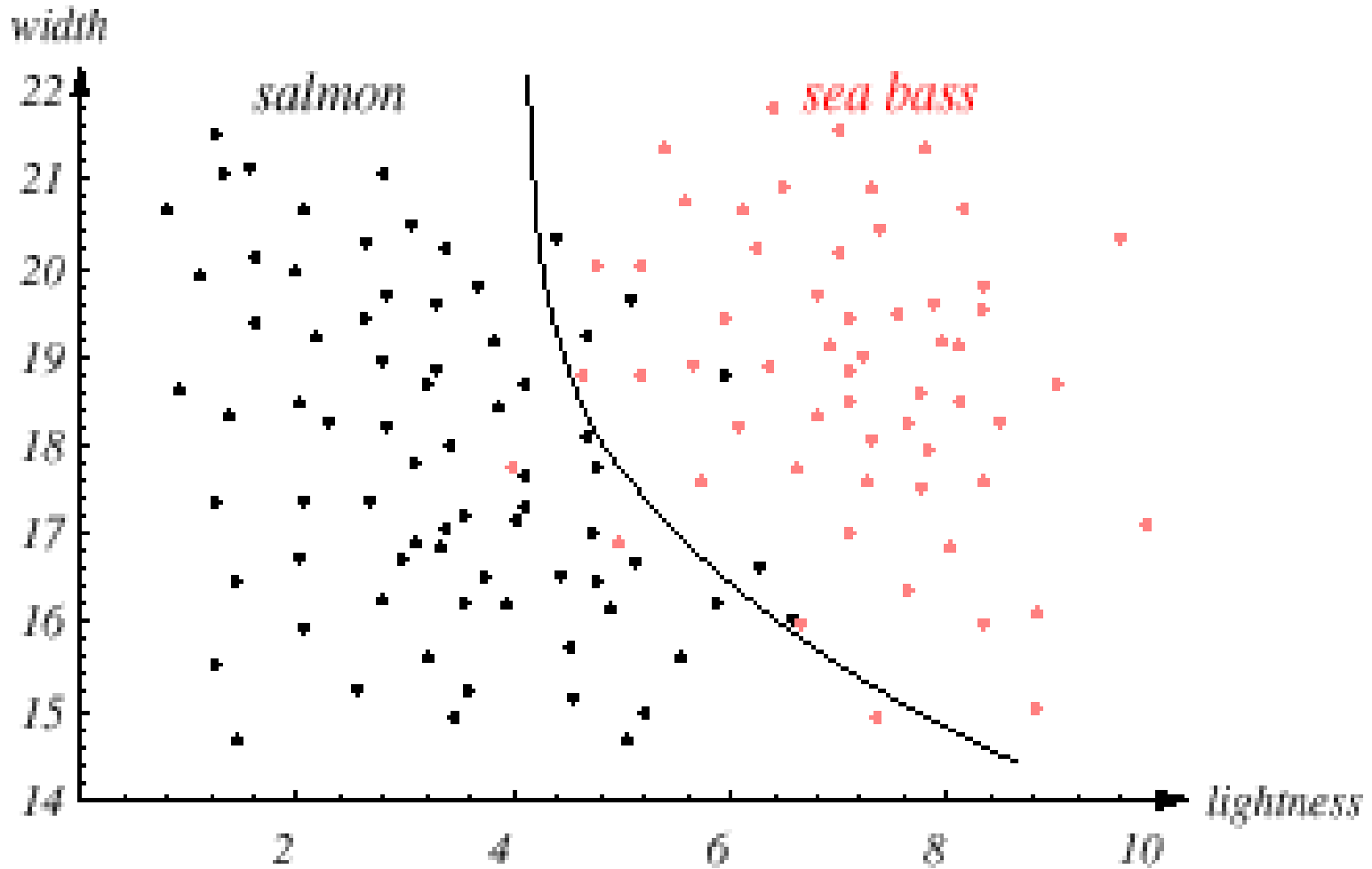
Decision/classification Boundaries

- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input



Issue of generalization!

Decision/classification Boundaries



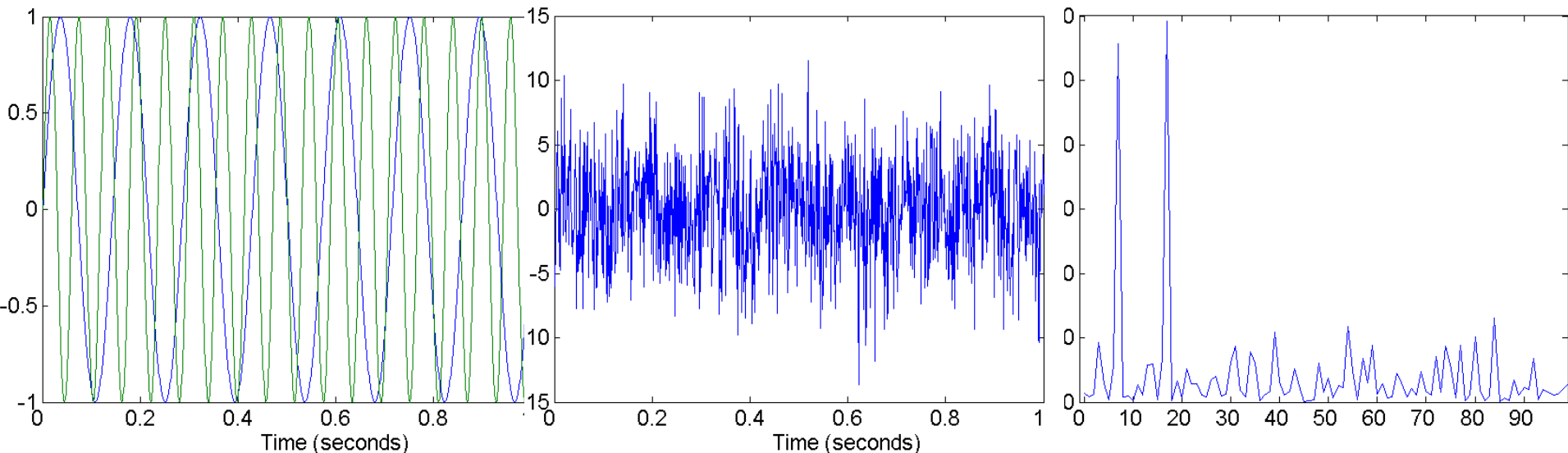
Occam's Razor

Entities are not to be multiplied without necessity

William of Occam
(1284-1347)

Mapping Data to a New Space

- Fourier transform
- Wavelet transform



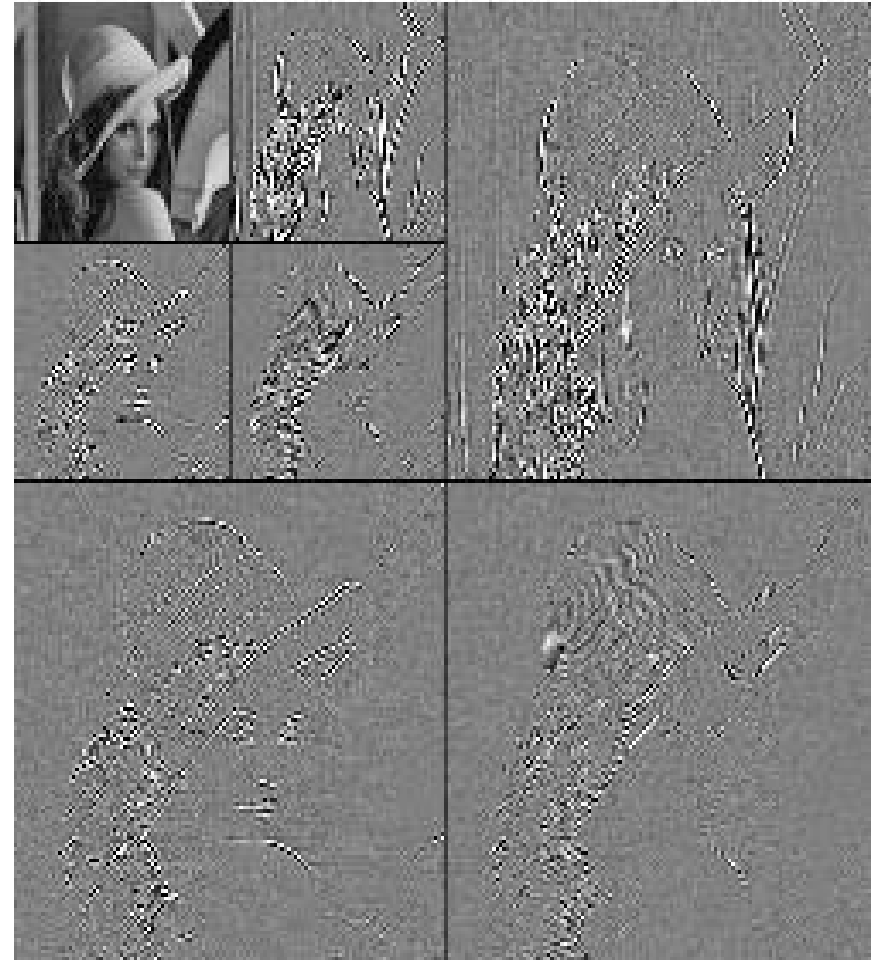
Two Sine Waves

Two Sine Waves + Noise

Frequency

What Is Wavelet Transform?

- Decomposes a signal into different frequency subbands
 - Applicable to n-dimensional signals
- Data are transformed to preserve relative distance between objects at different levels of resolution
- Allow natural clusters to become more distinguishable
- Used for image compression



What Is Wavelet Transform?

- Discrete wavelet transform (DWT) for linear signal processing, multi-resolution analysis
- Compressed approximation: store only a small fraction of the strongest of the wavelet coefficients
- Similar to discrete Fourier transform (DFT), but better lossy compression, localized in space
- Method:
 - Length, L , must be an integer power of 2 (padding with 0's, when necessary)
 - Each transform has 2 functions: smoothing, difference
 - Applies to pairs of data, resulting in two set of data of length $L/2$
 - Applies two functions recursively, until reaches the desired length

What Decomposition

- Wavelets: A math tool for space-efficient hierarchical decomposition of functions
- $S = [2, 2, 0, 2, 3, 5, 4, 4]$ can be transformed to $S_{\wedge} = [2^{3/4}, -1^{1/4}, 1/2, 0, 0, -1, -1, 0]$
- Compression: many small detail coefficients can be replaced by 0's, and only the significant coefficients are retained

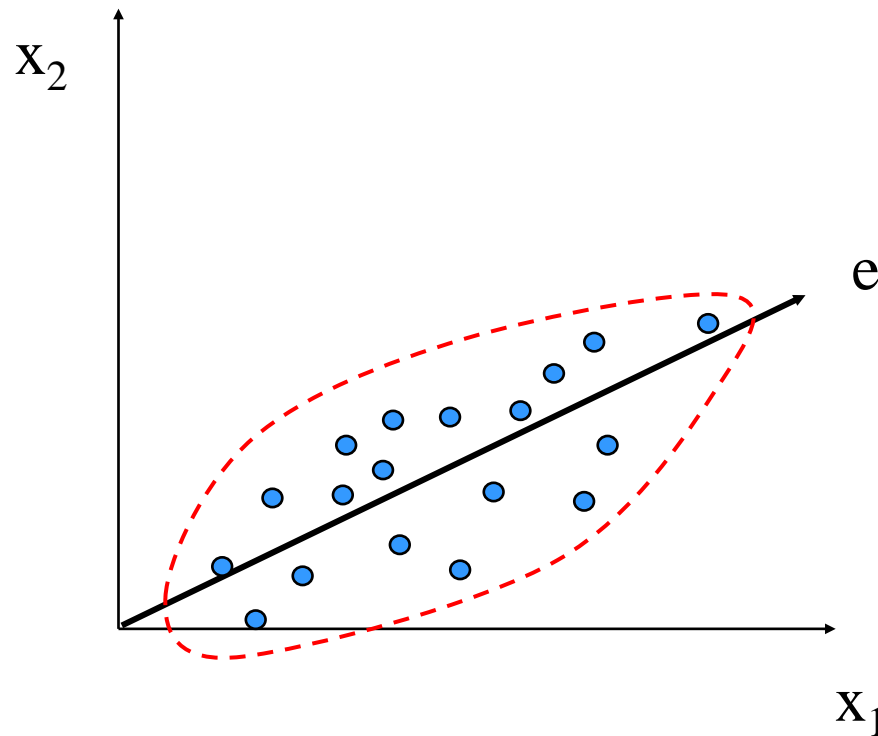
Resolution	Averages	Detail Coefficients
8	$[2, 2, 0, 2, 3, 5, 4, 4]$	
4	$[2, 1, 4, 4]$	$[0, -1, -1, 0]$
2	$[1\frac{1}{2}, 4]$	$[\frac{1}{2}, 0]$
1	$[2\frac{3}{4}]$	$[-1\frac{1}{4}]$

Why Wavelet Transform?

- Use hat-shape filters
 - Emphasize region where points cluster
 - Suppress weaker information in their boundaries
- Effective removal of outliers
 - Insensitive to noise, insensitive to input order
- Multi-resolution
 - Detect arbitrary shaped clusters at different scales
- Efficient
 - Complexity $O(N)$

Principal Component Analysis (PCA)

- Find a projection that captures the largest amount of variation in data
- The original data are projected onto a much smaller space, resulting in dimensionality reduction. We find the eigenvectors of the covariance matrix, and these eigenvectors define the new space



Principal Component Analysis (Steps)

- Given N data vectors from n -dimensions, find $k \leq n$ orthogonal vectors (*principal components*) that can be best used to represent data
 - Normalize input data: Each attribute falls within the same range
 - Compute k orthonormal (unit) vectors, i.e., *principal components*
 - Each input data (vector) is a linear combination of the k principal component vectors
 - The principal components are sorted in order of decreasing “significance” or strength
 - Since the components are sorted, the size of the data can be reduced by eliminating the *weak components*, i.e., those with low variance (i.e., using the strongest principal components, it is possible to reconstruct a good approximation of the original data)
- Works for numeric data only

Feature Subset Selection: revisited

- Another way to reduce dimensionality of data
- Redundant attributes
 - Duplicate much or all of the information contained in one or more other attributes
 - E.g., purchase price of a product and the amount of sales tax paid
- Irrelevant attributes
 - Contain no information that is useful for the data mining task at hand
 - E.g., students' ID is often irrelevant to the task of predicting students' GPA

Feature Subset Selection

- **Techniques:**
 - Brute-force approach:
 - Try all possible feature subsets as input to data mining algorithm
 - Embedded approaches:
 - Feature selection occurs naturally as part of the data mining algorithm
 - Filter approaches:
 - Features are selected before data mining algorithm is run
 - Wrapper approaches:
 - Use the data mining algorithm as a black box to find best subset of attributes

Example:
Classification and
Clustering

Validation of models

Semisupervised and Co-Training

Multiple Classification Systems

Granular Computing

GrC practical

Pattern Recognition System (PRS)

Measurement → *Feature* → *Decision*
Space *Space* *Space*

- Uncertainties arise from deficiencies of information available from a situation
- Deficiencies may result from incomplete, imprecise, ill-defined, not fully reliable, vague, contradictory information in various stages of a PRS