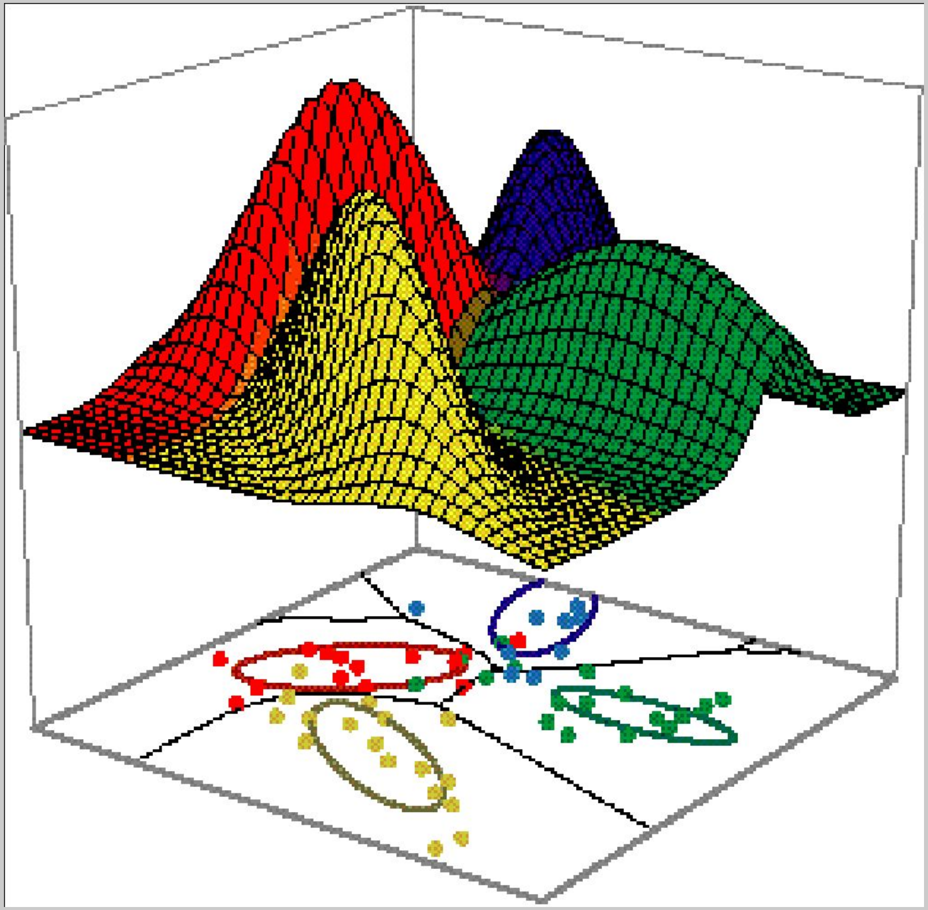


Pattern Recognition (Αναγνώριση Προτύπων)

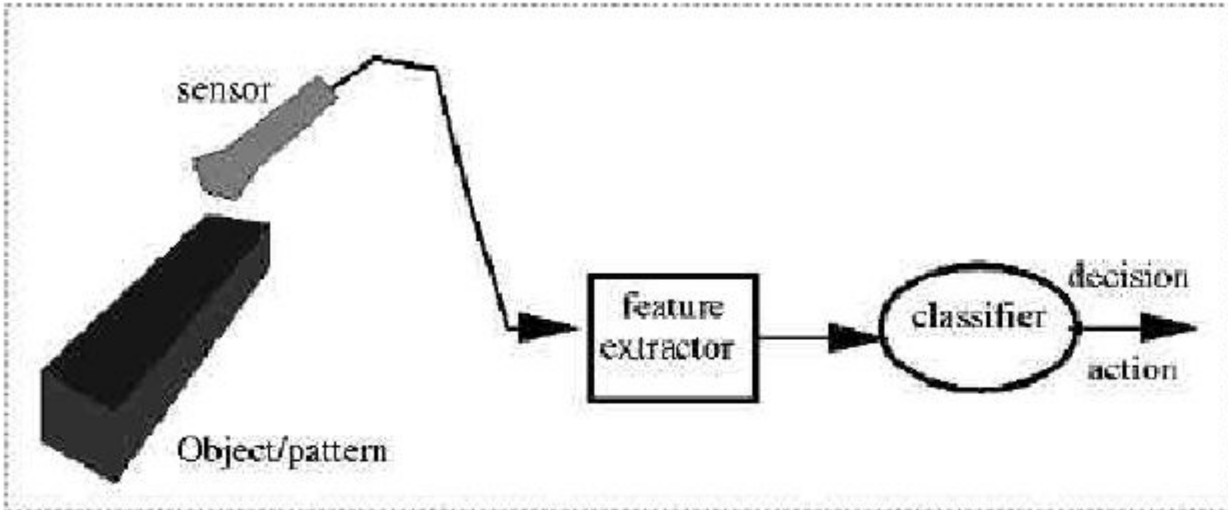
Panos Trahanias

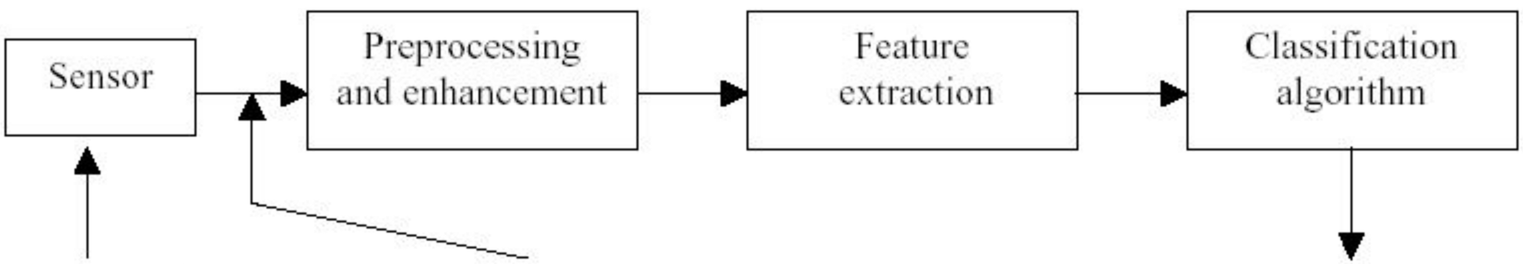
UNIVERSITY OF CRETE
DEPARTMENT of
COMPUTER SCIENCE

Section 1: Introduction to Learning Theory

- **Pattern Recognition**: The science / technology that automatically detects & recognizes useful regularities in noisy and complex environments.
- Extremely wide! No theory alone fully covers all aspects.
- Classification Theory (Θεωρία Ταξινόμησης): The process of classifying/distinguishing “things” into different groups/classes.
- It is based on statistical decision theory and provides mathematical procedures for representing features in the form of vectors.

Standard System

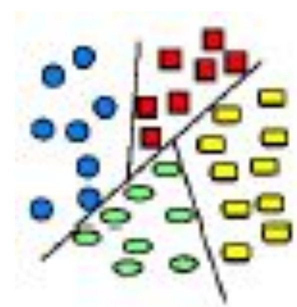




Observed world: pattern data

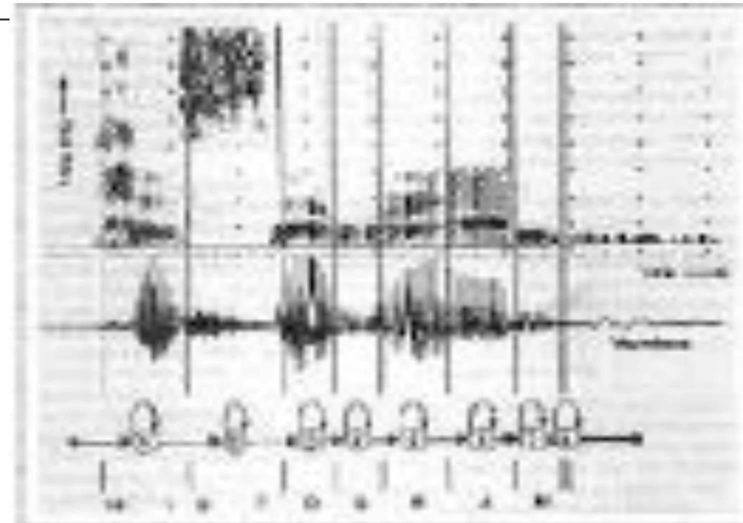
Measurements

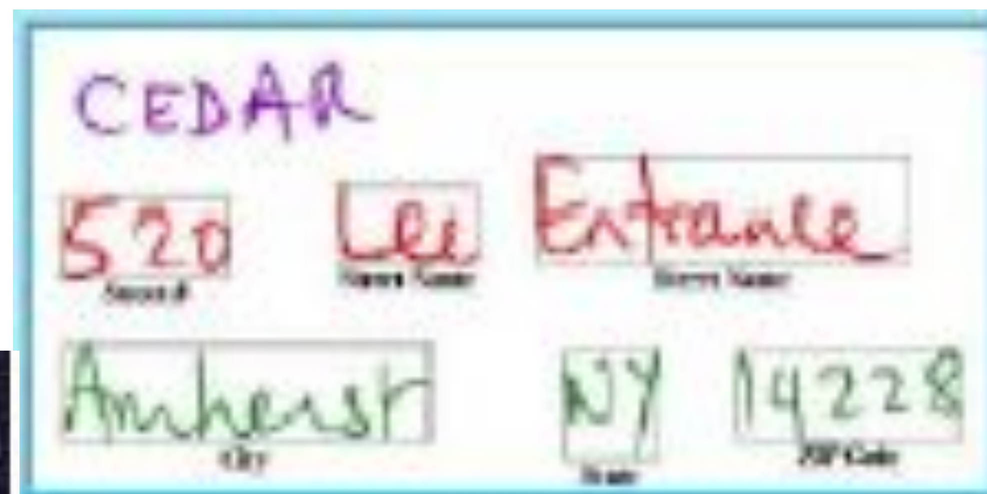
Classification



Machine Perception (Μηχανική Αντίληψη)

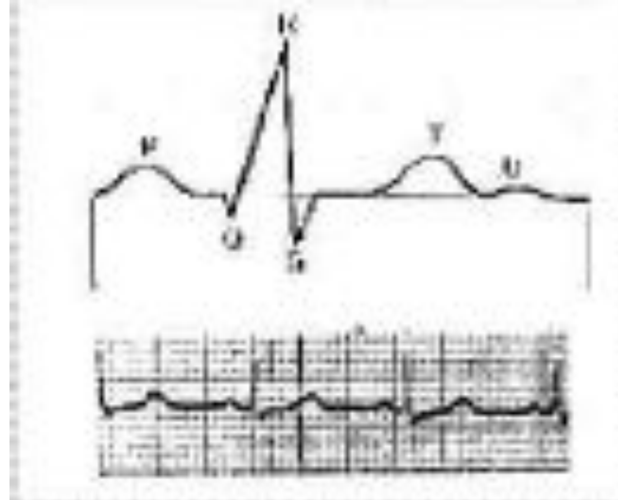
- Implementation of a machine (program) that can recognize and classify patterns
- Speech recognition / speaker identification
- Handwritten character recognition
- Optical character recognition - OCR
- Fingerprint identification

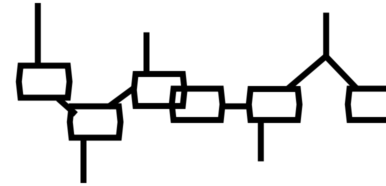


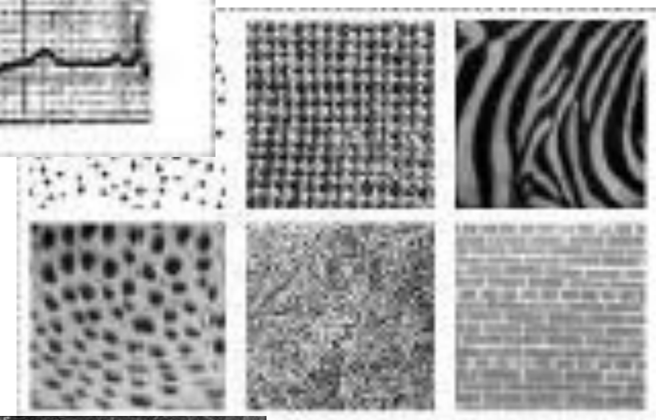


Applications

- DNA sequence identification
- X-Ray Mammography test
- EEG - Neurological disorder diagnosis
- ECG - Cardiovascular disorder diagnosis
- Iris scan identification
- Topographical remote sensing

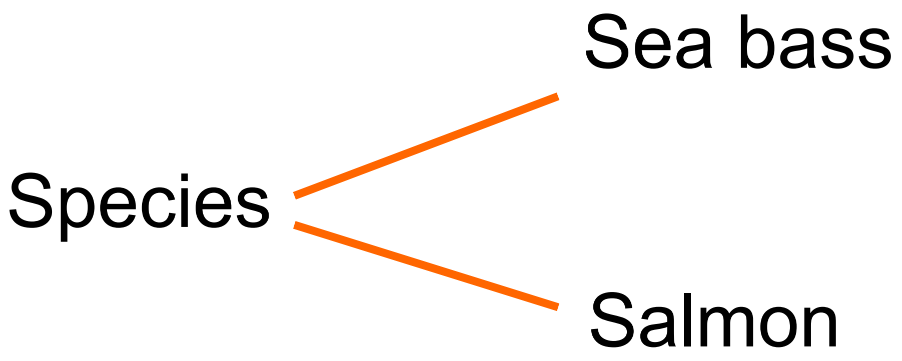






Example

- “Classify fish on a conveyor belt by species, using an optical sensor”



Problem Analysis

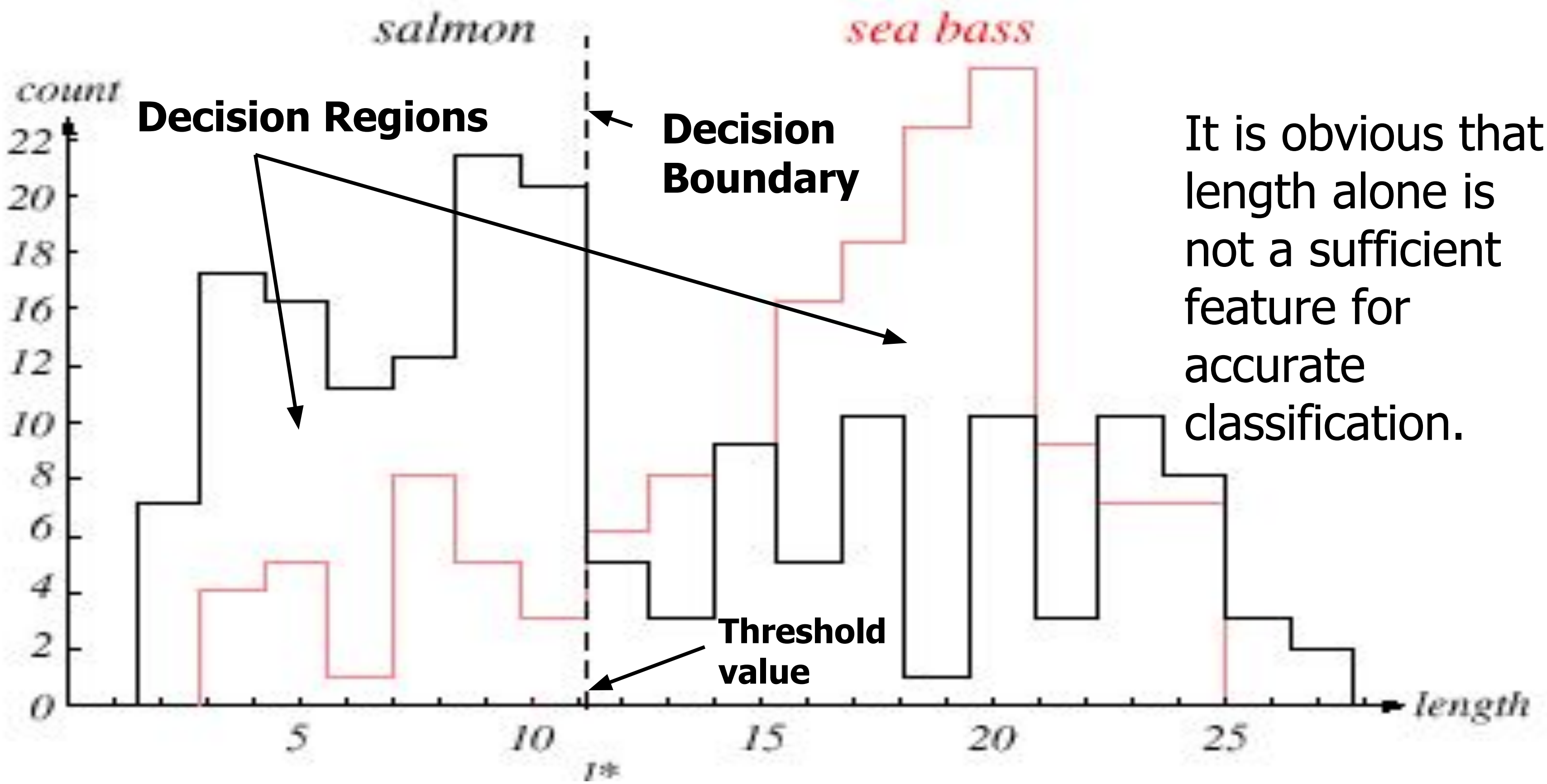
- We place the camera to capture a certain number of images. The images are sent to the feature extractor. The <u>feature extractor</u> reduces the data to a small number of <u>features</u>.
  - Length
  - Brightness
  - Width
  - Number and shape of the fins
  - Position of the mouth, etc.
- This is a small set of properties whose values are given as inputs to the classifier.

Classification System

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.

Length Histogram

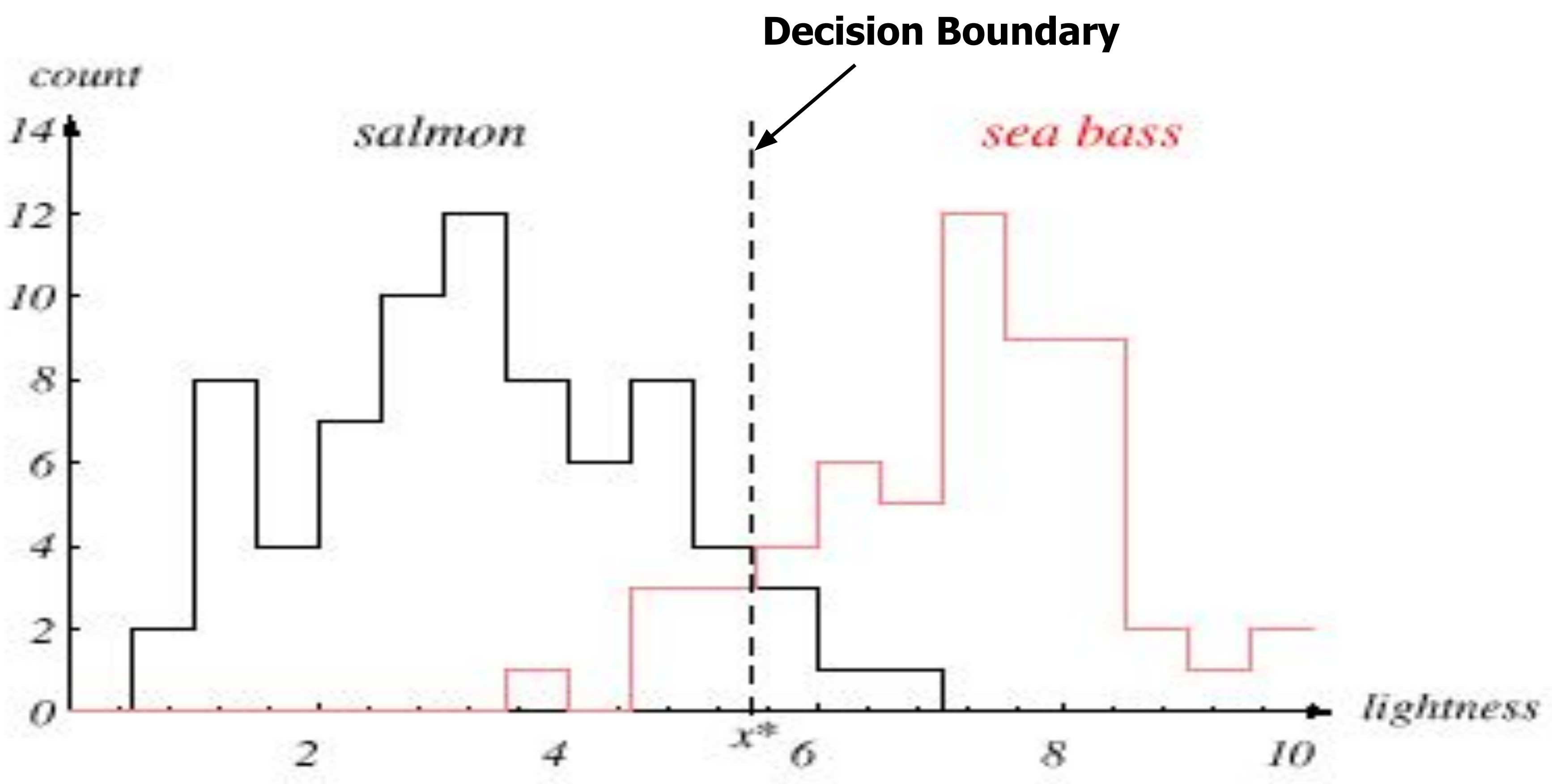
- Select length as the unique feature for the classification:



It is obvious that length alone is not a sufficient feature for accurate classification.

Length histograms for both categories. There is no critical length value that can clearly distinguish between the two categories. Using length only creates errors. The value l* creates the smallest number of errors on average.

Brightness Histogram



Brightness histograms for both categories. As in the case of length, there is no critical brightness value that can clearly distinguish between the two categories. Value X* gives rise to the smallest number of errors on average.

Concept of Cost

- The relationship between the selection of the critical value (threshold) and the cost of incorrect classification is obvious.

- In the real world, different errors are associated with different costs.

- In the problem of separating salmon from sea bass, the cost of incorrectly classifying sea bass as salmon is higher than that of incorrectly classifying salmon as sea bass!

- Therefore move the decision limit to lower values (length or brightness), in order to reduce the number of sea bass that are incorrectly classified as salmon !!

- This is the work of Statistical Decision Theory.

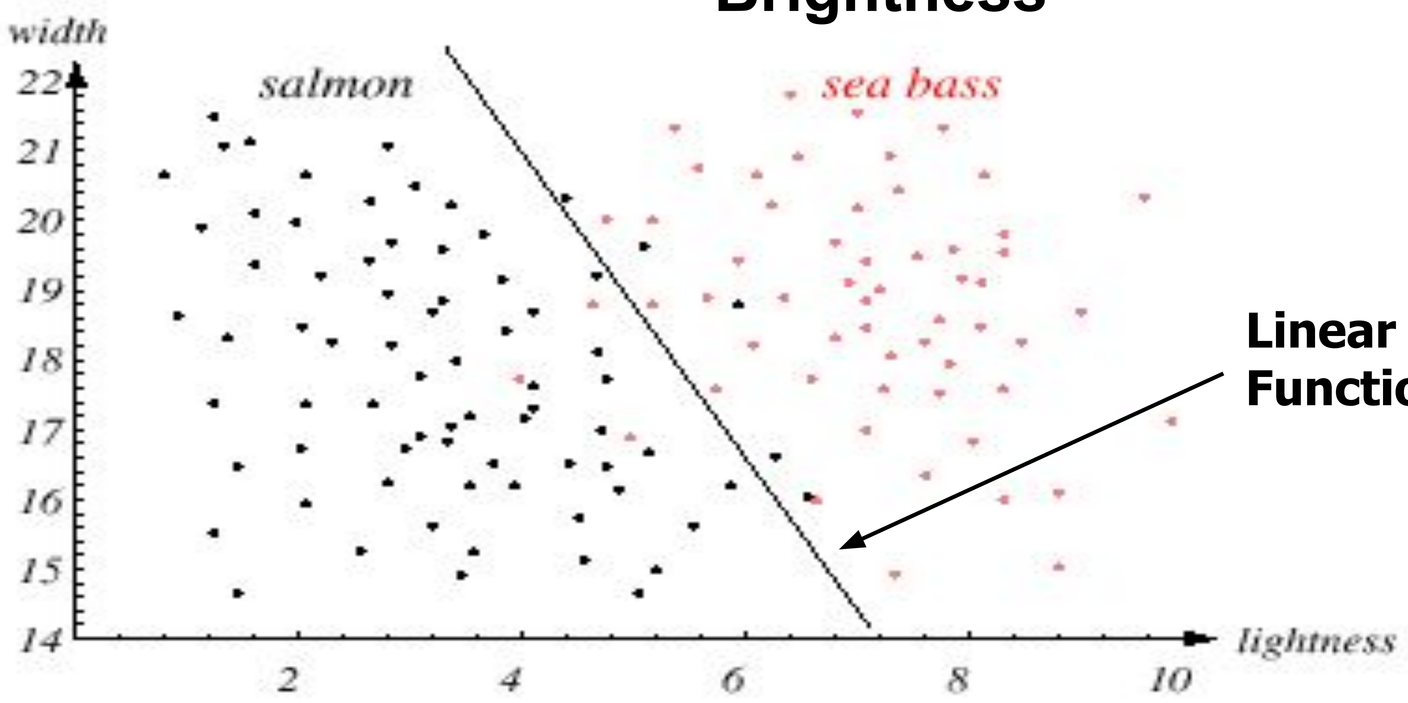
Feature Vector

- Use of brightness and length for classification:

Fish → $x^T = [x_1, x_2]$

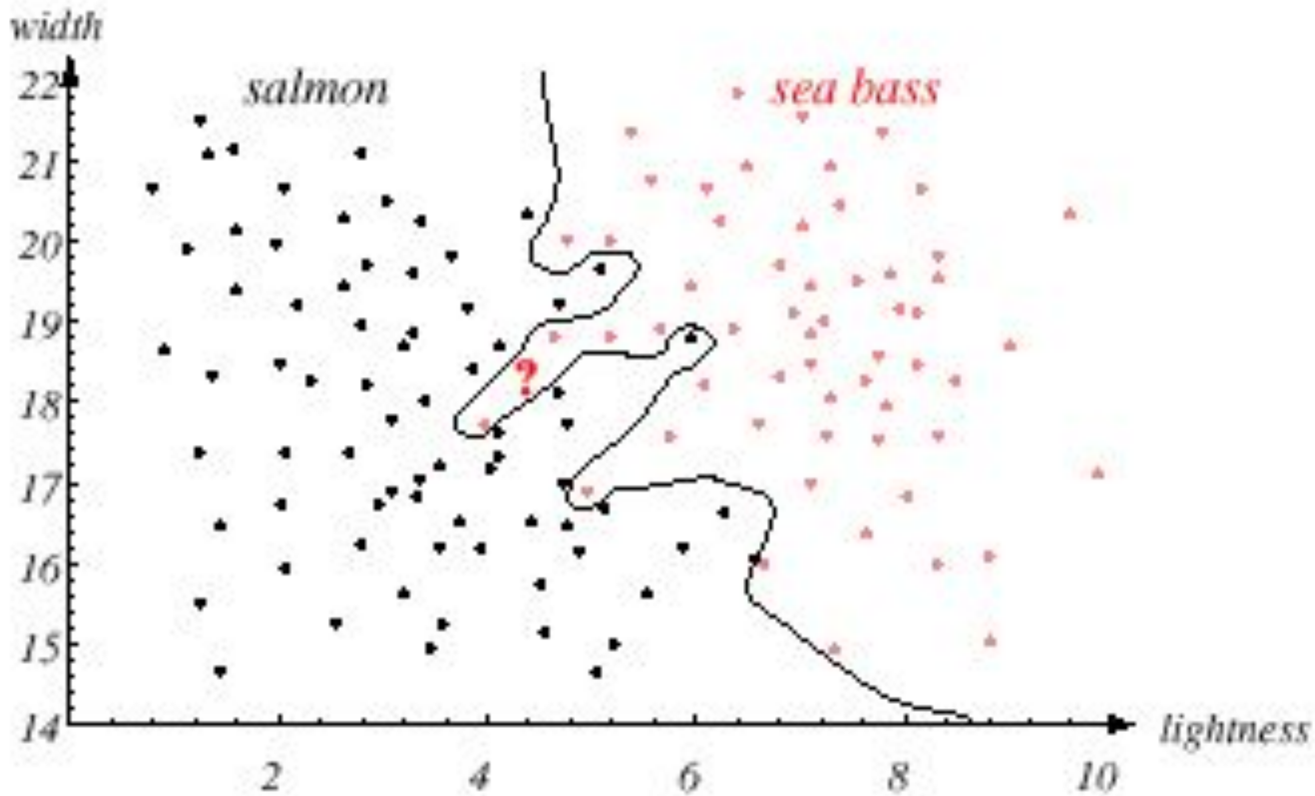
Brightness

Lengt h



Linear Discriminant Function

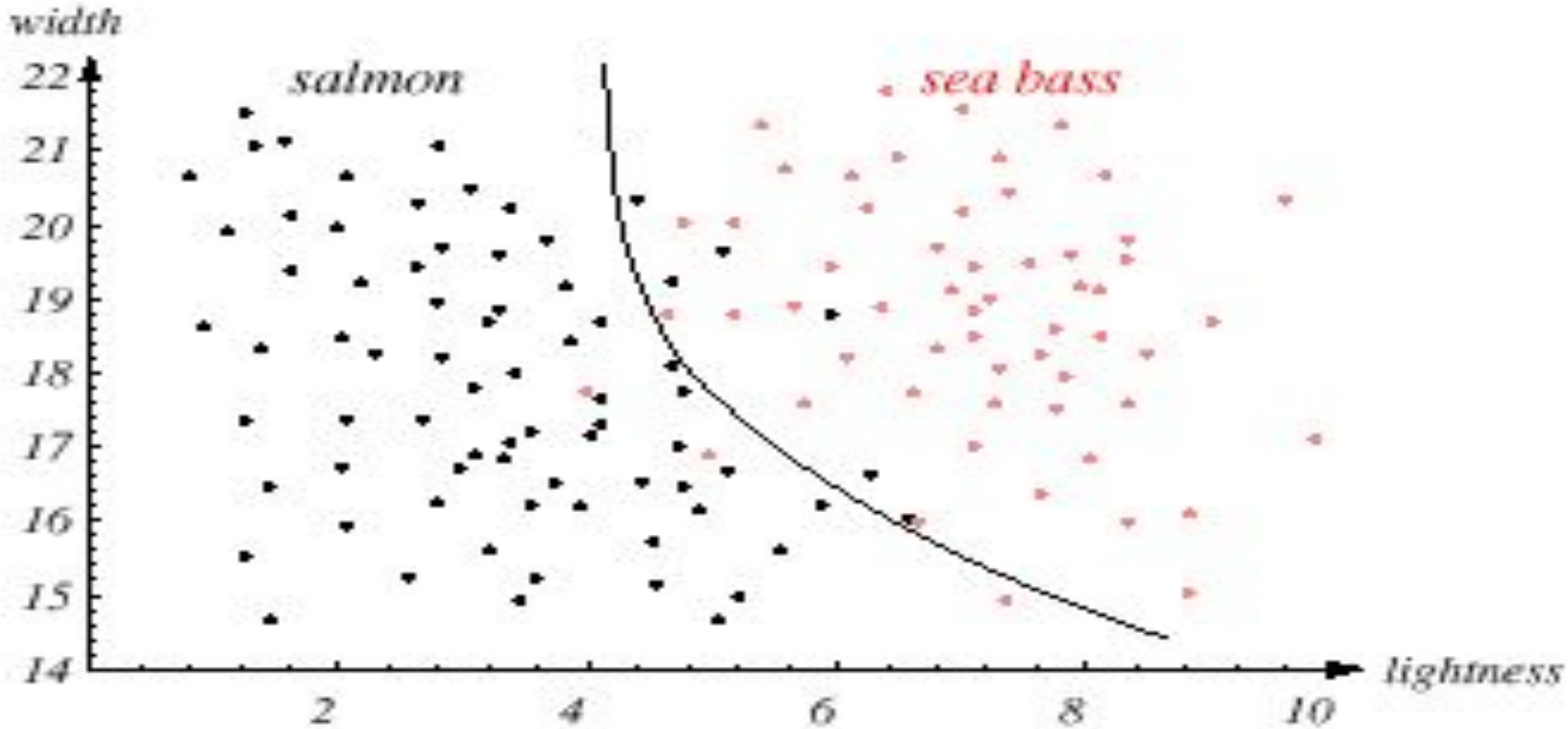
Nonlinear Functions



Complex models for fish lead to complex decision boundaries: Such a choice leads to a perfect classification of samples with known class (training the classifier) but can lead to poor classification of new patterns. For example, the new point **?** is obviously salmon with high probability, but the discrimination function leads to its classification as sea bass.

Generalization of Classifier Design

The main goal is to classify <u>new</u> input samples



The discriminant function must reconcile the need for the best possible classification of system training samples with the need for simple classifiers. Following this principle we achieve the greatest possible accuracy in the classification of new samples.

Pattern Recognition System - PRS

Sensor

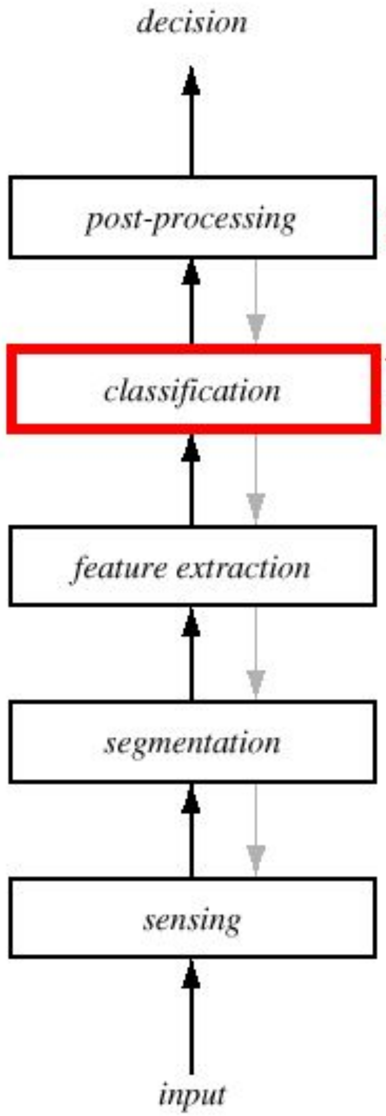
- Use of transducer sensor (camera, microphone).
- PRS depends on the bandwidth, resolution, signal distortions of the transducer.

Segmentation and grouping

- Objects must be well separated, without overlapping.

Feature extraction

- "Insightful" features.
- Characteristics unchanged in scaling, turning, reflection.



Classification

- Use of feature vector to classify the object into a category.

Post-processing

- Exploit other relevant information to improve performance.

Design Stages

<u>Data Collection</u>

- Sufficient data for adequate training, testing and evaluation of the system.

<u>Feature Choice</u>

- How to choose them, noise effect, simplicity, etc.

<u>Model Choice</u>

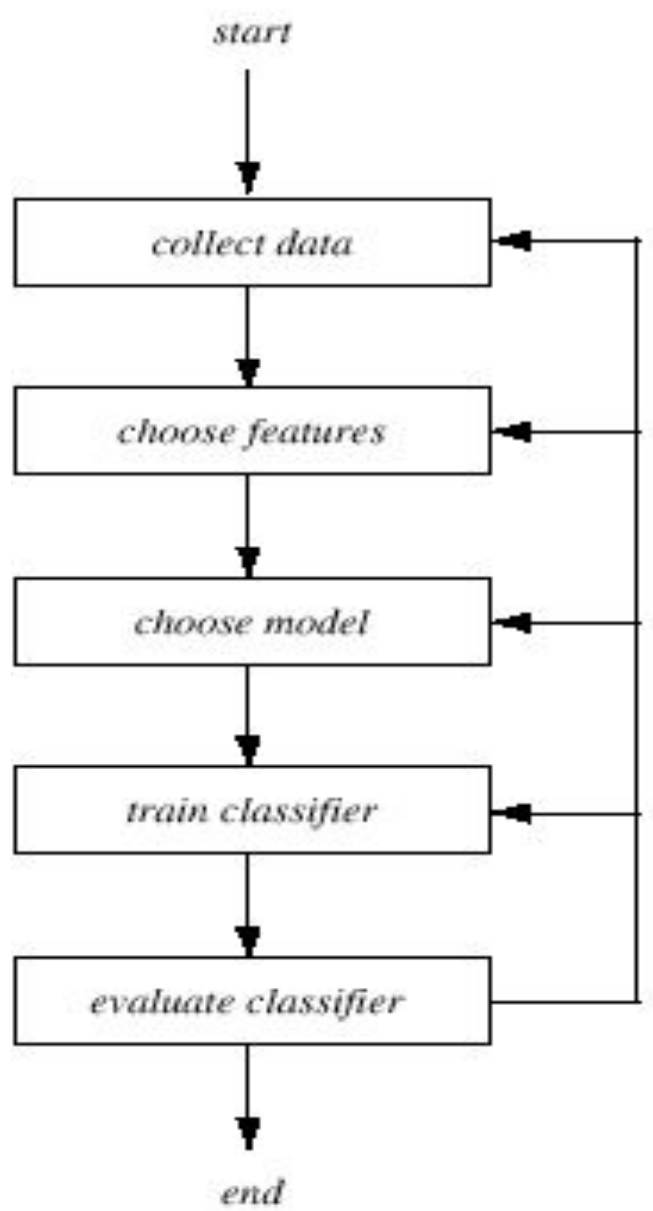
- Probability Density Functions.

<u>Training</u>

- Methodology of training the classifier from the data.

<u>Evaluation</u>

- Measurement of classification error, change from one feature set to another.
- Computational Complexity: Performance as a function of design parameters.



Learning and Adaptation

Supervised Learning

- It assumes that the samples used in the learning phase have a label indicating the class to which they belong.

Unsupervised Learning

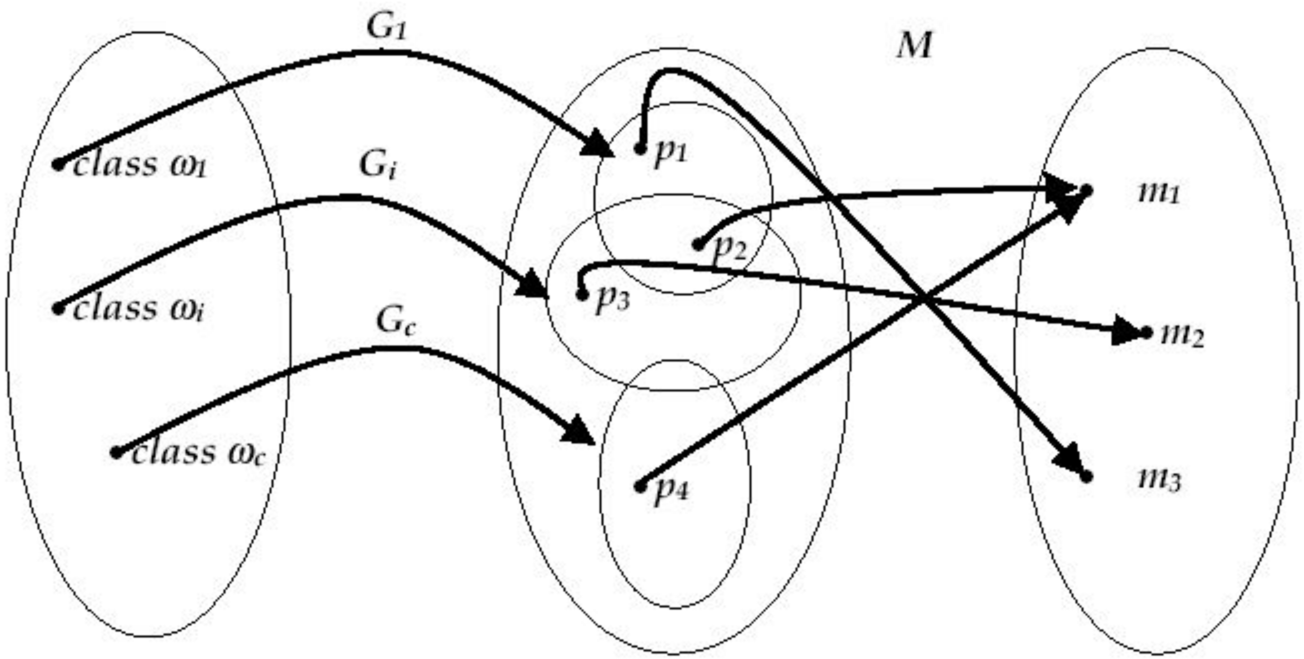
- It is based on a collection of samples for which we are unaware of their classification. The system forms clusters of input patterns.

Adaptation

- The characteristics of the patterns change over time and therefore the classifier must follow the changes without supervision.

Pattern Recognition

<u>Accordingly</u>: Pattern Recognition can be characterized as a process of <u>information reduction</u>, <u>information mapping</u>, or <u>information labeling</u>.



Class Membership – Description Space, *Ω*

Realized Pattern Space, *P*

Measurement Space, *F*

Mathematical Background

- Linear Algebra
  - Determinants and Inversion of Tables
  - Eigenvalues and Eigenvectors
  - Differentiation of Tables
- Probability theory
  - Discrete random variables (r.v.)
  - Pairs of r.v.
  - Mean values, standard deviations, variability
  - Statistical independence / correlation
  - Conditional Probability
  - Law of total probability and Bayes' Rule
  - Normal distribution and central limit theorem

Course Syllabus

- **Introduction to Learning Theory**
  - Learning machines
  - Pattern recognition systems
  - The system design cycle
  - Learning and adaptability
- **Bayesian Decision Theory**
  - Classifiers, discrimination functions, decision surfaces
  - Minimum probability of error Classification
  - (Gaussian) Bayes classifier for binary classification
  - Classification error probability
  - Bayesian belief networks
- **Parameter Estimation and Supervised Learning**
  - Maximum likelihood estimation
  - Bayesian estimation
- **Non-Parametric Techniques**
  - Distribution estimation and Parzen windows
  - k-Nearest neighbors
  - Methodologies for reducing the number of patterns
- **Linear Discrimination Functions**
  - Linearly separable classes
  - Perceptron Algorithm
  - Least squares method
- **Artificial Neural Networks (ANN)**
  - Introduction and structure
  - Recursive ANN
  - Back-propagation algorithm
  - Multilayer Perceptron
- **Stochastic Methods and Unsupervised Learning**
  - Sample density and recognizability
  - Recursive ANN
  - Unsupervised learning with Gaussian sample density
  - Data grouping