Lecture 1: Course introduction

- **Course organization**
  - Grading policy
  - Outline and calendar
- **Introduction to pattern recognition**
  - Definitions and related terms
  - Features and patterns
  - Decision regions and discriminant functions
- **Pattern recognition examples**
- **Pattern recognition approaches**
  - Statistical
  - Neural
  - Structural

Course organization

- **Instructor**
  - Ricardo Gutierrez-Osuna
  - Office: 401 RC
  - Tel: (937) 775-5120
  - Email: **`rgutier@cs.wright.edu`**
  - **`http://www.cs.wright.edu/~rgutier`**
  - Office hours: MW 4:00-5:30 PM
- **Meeting time and location**
  - 302 Russ Engineering Center, MW 5:35-6:50 PM
- **Grading**
  - Homework
    - Bi-weekly (first 6 weeks)
  - Exams
    - 1 midterm and 1 final
  - Term Project
    - Open-ended
    - In-class presentation

	Weight (%)
Homework	30
Project	30
Midterm	20
Final Exam	20

Course outline

- **Introduction to Pattern Recognition (1)**
  - What is pattern recognition?
  - Approaches to pattern recognition: statistical, neural and structural
- **Overview of Background Material (2)**
  - Random variables and Probability
  - Linear Algebra
  - MATLAB®
- **Decision Theory (1)**
  - Likelihood Ratio Test
  - Probability of error, Bayes Risk
- **Dimensionality reduction (2)**
  - The curse of dimensionality
  - Principal Components Analysis
  - Linear Discriminant Analysis
- **Statistical Classifiers (2)**
  - Linear and quadratic classifiers
  - The K Nearest Neighbor (KNN) classifier
- **Density Estimation (2)**
  - Parameter estimation: Maximum Likelihood
  - Non-Parametric density estimation: Histograms, Kernels, KNN
  - Optimal and Naïve Bayes Classifiers
- **Unsupervised Learning (2)**
  - K-means and ISODATA
  - Hierarchical clustering
  - Competitive Learning
  - Kohonen Self-Organizing Maps
- **Feature Selection (2)**
  - Search strategies: exhaustive, sequential, randomized
  - Evaluation strategies: filter, wrapper
- **Validation (1)**
  - Holdout, cross-validation, bootstrap
  - Data splits
- **Classification using Multilayer Perceptrons (2)**
  - Historical overview
  - Learning: back-prop and enhancements
  - The role of hidden and output units

Tentative quarter calendar

	Date	Topic	Reading (chapters)	Assignments
Week 1	12/31	(No class)		
	1/2	Course introduction	1	
Week 2	1/7	Random variables, Probability	A.4, A.5	
	1/9	Linear Algebra, MATLAB®	A.2	
Week 3	1/14	Bayesian Decision Theory	2.1-3	HW1 assigned
	1/16	Dimensionality reduction: Principal Components Analysis	3.7, 3.8.1	
Week 4	1/21	Martin Luther King, Jr. Day (No class)		
	1/23	Dimensionality reduction: Linear Discriminants Analysis	3.8.2	
Week 5	1/28	Linear and quadratic classifiers	2.4-7	HW1 due HW2 assigned
	1/30	The K Nearest Neighbors classifier	4.5-6	
Week 6	2/4	Midterm		
	2/6	Parameter estimation, histograms, KNN	4.1-2, 4.4	
Week 7	2/11	Kernel Density Estimation	4.3	HW2 due HW3 assigned
	2/13	Unsupervised learning: statistical clustering	10.6-9	
Week 8	2/18	Unsupervised learning: Competitive Learning, Kohonen SOM	10.11, 10.14	
	2/20	Feature selection I: objective functions, sequential FS		Project proposal due
Week 9	2/25	Feature selection II: exponential and randomized FS	7.2.1-2 7.5-6	HW3 due
	2/27	Validation	9.1-2, 9.4 9.6.1-3	
Week 10	3/4	Multi-layer perceptrons: history, back-prop, enhancements	6.1-4	
	3/6	Multi-layer perceptrons: the role of hidden and output units	6.5-8	
Week 11	3/11	Final Exam		
	3/13	Project Presentations 5:30-7:30 PM, RC 302		Project report due

Definition of pattern recognition

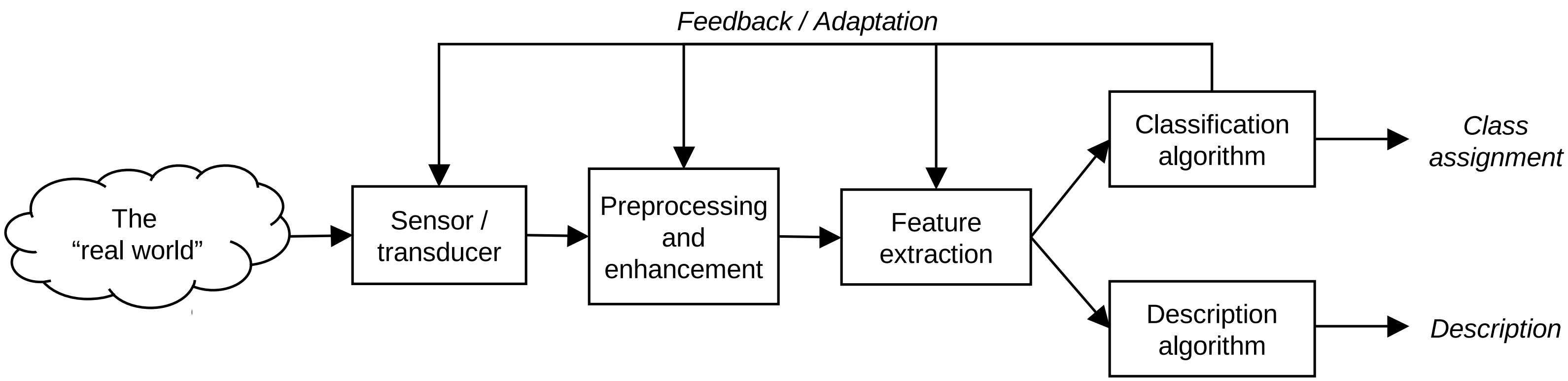
- **Definitions from the literature**
  - “*The assignment of a physical object or event to one of several pre-specified categories*” --Duda & Hart
  - “*A problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories or classes*” --Fukunaga
  - “Given some examples of complex signals and the correct decisions for them, make decisions automatically for a stream of future examples” –Ripley
  - *“The science that concerns the description or classification (recognition) of measurements”* --Schalkoff
  - “*The process of giving names ω to observations **x***”, --Schürmann
  - Pattern Recognition is concerned with answering the question ***“What is this?”*** --Morse

More on Pattern Recognition

- **Related fields**
  - Adaptive Signal Processing
  - Machine Learning
  - Artificial Neural Networks
  - Robotics and Vision
  - Cognitive Sciences
  - Mathematical Statistics
  - Nonlinear Optimization
  - Exploratory Data Analysis
  - Fuzzy and Genetic systems
  - Detection and Estimation Theory
  - Formal Languages
  - Structural Modeling
- **Applications**
  - Image Preprocessing / Segmentation
  - Computer Vision
  - Speech Recognition
  - Automated Target Recognition
  - Optical Character Recognition
  - Seismic Analysis
  - Man and Machine Diagnostics
  - Fingerprint Identification
  - Industrial Inspection
  - Financial Forecast
  - Medical Diagnosis
  - EKG Signal Analysis

Components of a pattern recognition system

- **A typical pattern recognition system contains**
  - A sensor
  - A preprocessing mechanism
  - A feature extraction mechanism (manual or automated)
  - A classification or description algorithm
  - A set of examples (training set) already classified or described



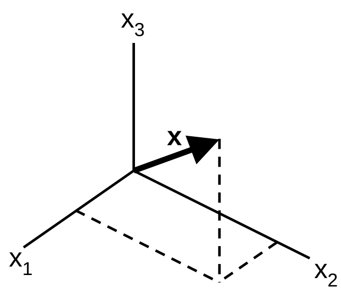
Features and patterns (1)

Feature

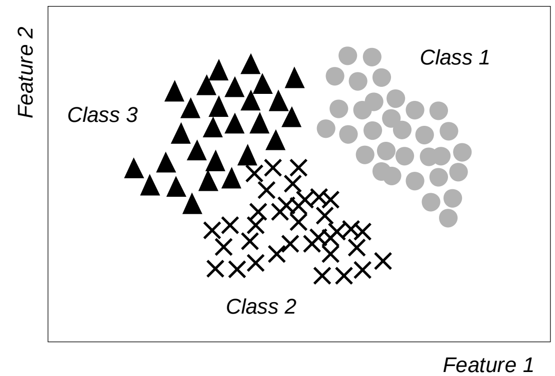
- Feature is any <u>distinctive</u> aspect, quality or characteristic
  - Features may be symbolic (i.e., color) or numeric (i.e., height)
- Definitions
  - The combination of d features is represented as a d-dimensional column vector called a **feature vector**
  - The d-dimensional space defined by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

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

Feature vector



Feature space (3D)



Scatter plot (2D)

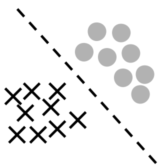
Pattern

- Pattern is a <u>composite</u> of traits or features <u>characteristic of an individual</u>
- For our purposes, a pattern is a <u>pair</u> of variables $\{x, \omega\}$ where
  - ***x*** is a collection of observations or features (feature vector)
  - ω is the concept behind the observation (label)

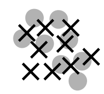
Features and patterns (2)

What makes a "good" feature vector?

- 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 different classes have different feature values

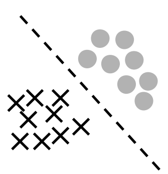


"Good" features

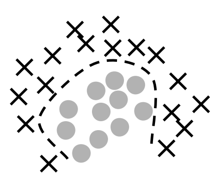


"Bad" features

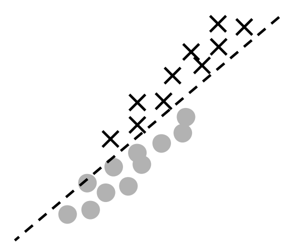
More feature properties



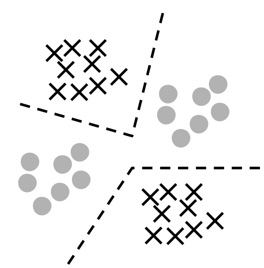
Linear separability



Non-linear separability



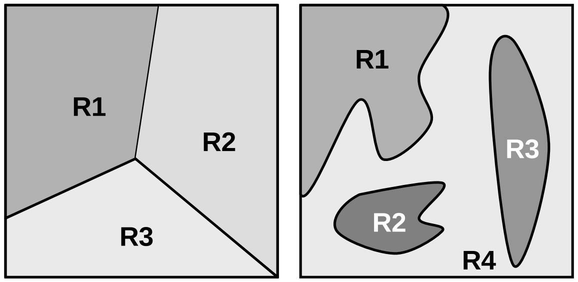
Highly correlated features



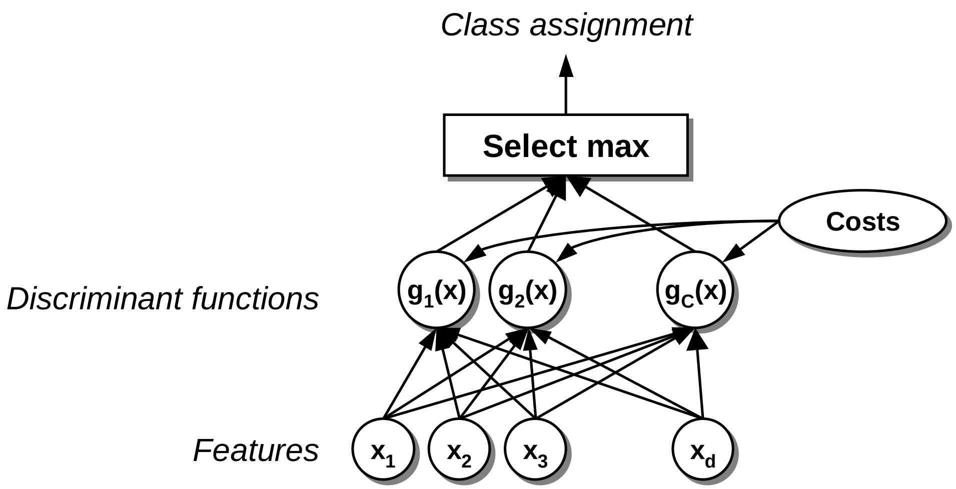
Multi-modal

Classifiers

- **The task of a classifier is to partition feature space into class-labeled decision regions**
  - Borders between decision regions are called **decision boundaries**
  - The classification of feature vector ***x*** consists of determining which decision region it belongs to, and assign ***x*** to this class



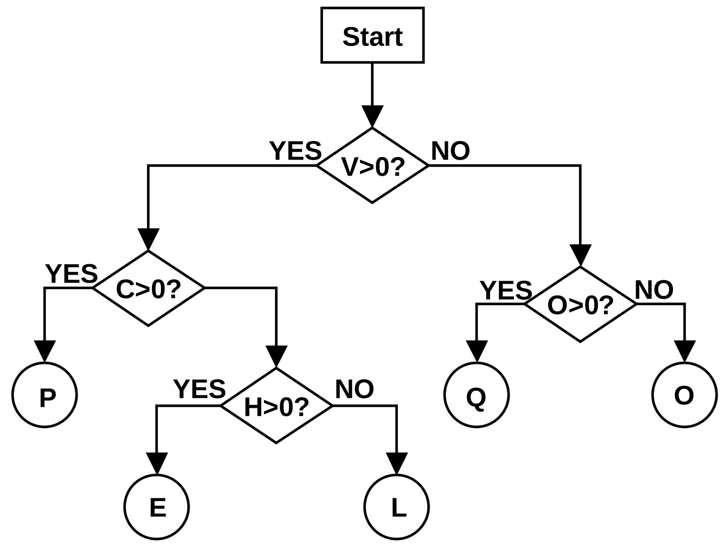
- **A classifier can be represented as a set of discriminant functions**
  - The classifier assigns a feature vector ***x*** to class ω_i if $g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}) \quad \forall j \neq i$



A simple pattern recognition problem

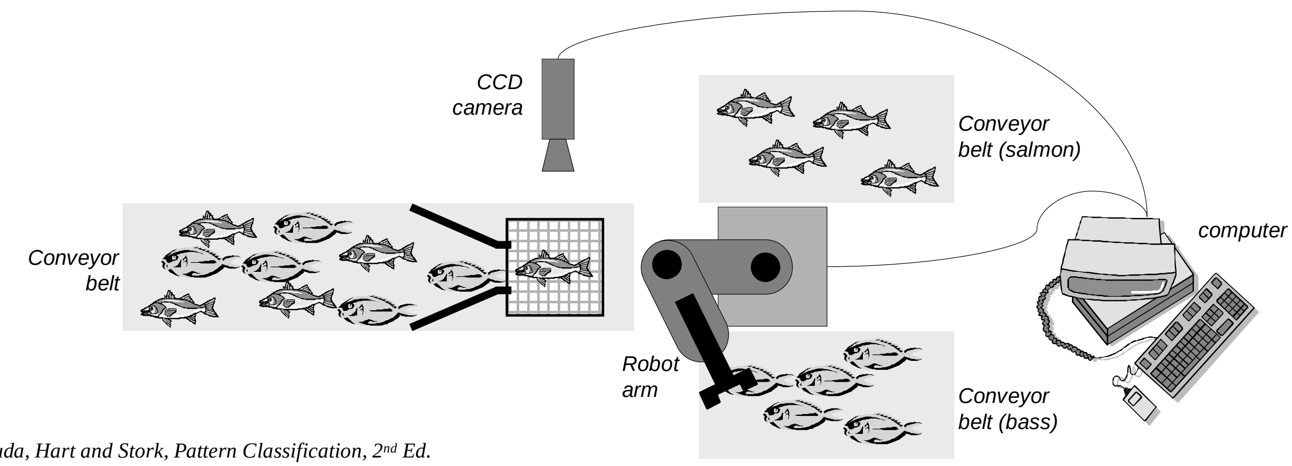
- **Consider the problem of recognizing the letters L,P,O,E,Q**
  - Determine a sufficient set of features
  - Design a tree-structured classifier

Character	Features			
	Vertical straight lines	*Horizontal straight lines*	*Oblique straight lines*	*Curved lines*
L	*1*	*1*	*0*	*0*
P	*1*	*0*	*0*	*1*
O	*0*	*0*	*0*	*1*
E	*1*	*3*	*0*	*0*
Q	*0*	*0*	*1*	*1*



A realistic pattern recognition system (1)

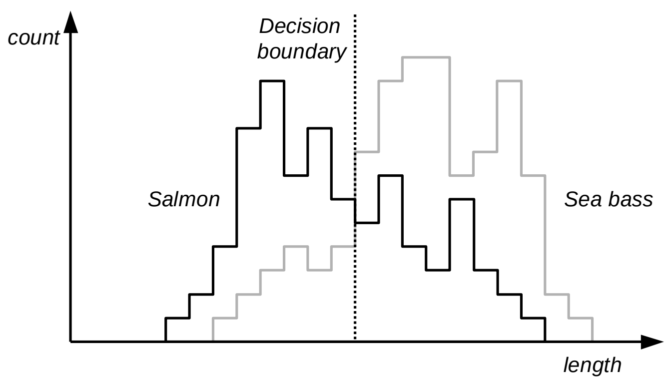
- **Consider the following scenario***
  - A fish processing plan wants to automate the process of sorting incoming fish according to species (salmon or sea bass)
  - The automation system consists of
    - a conveyor belt for incoming products
    - two conveyor belts for sorted products
    - a pick-and-place robotic arm
    - a vision system with an overhead CCD camera
    - a computer to analyze images and control the robot arm



**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

A realistic pattern recognition system (2)

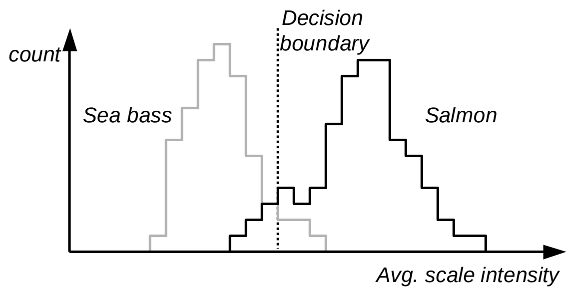
- **Sensor**
  - The vision system captures an image as a new fish enters the sorting area
- **Preprocessing**
  - Image processing algorithms
    - adjustments for average intensity levels
    - segmentation to separate fish from background
- **Feature Extraction**
  - Suppose we know that, on the average, sea bass is larger than salmon
    - From the segmented image we estimate the length of the fish
- **Classification**
  - Collect a set of examples from both species
  - Compute the distribution of lengths for both classes
  - Determine a decision boundary (threshold) that minimizes the classification error
  - We estimate the classifier's probability of error and obtain a discouraging result of 40%
  - **What do we do now?**



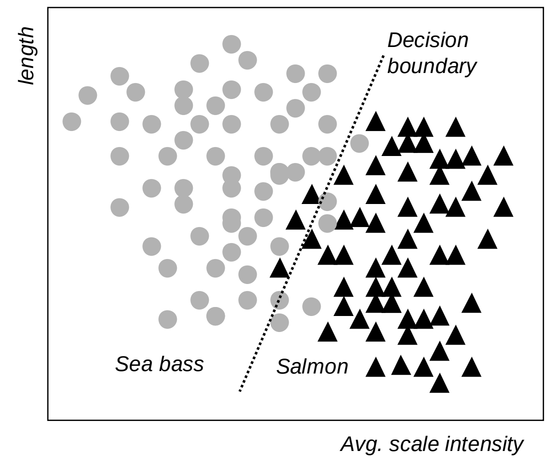
A realistic pattern recognition system (3)

■ Improving the performance of our PR system

- Determined to achieve a recognition rate of 95%, we try a number of features
  - Width, Area, Position of the eyes w.r.t. mouth...
  - only to find out that these features contain no discriminatory information
- Finally we find a "good" feature: average intensity of the scales

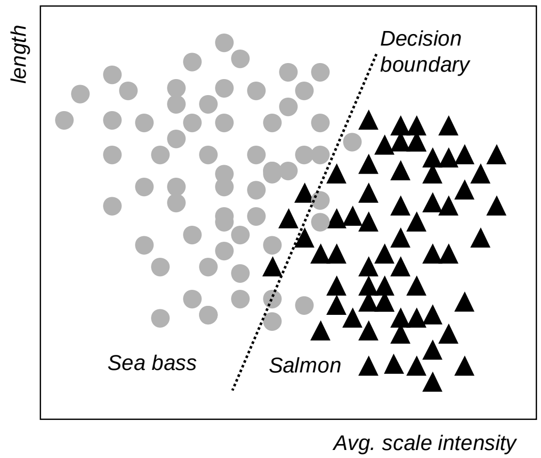


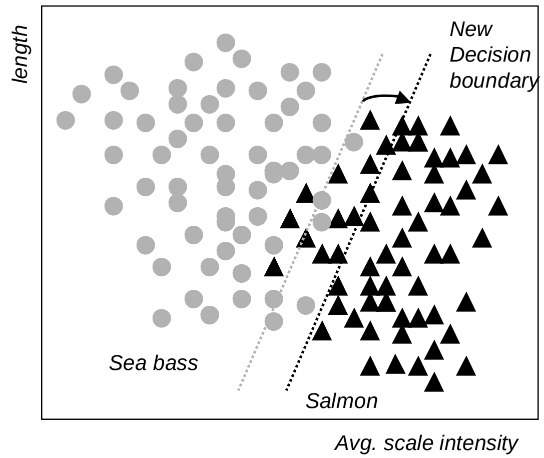
- We combine "*length*" and "*average intensity of the scales*" to improve class separability
- We compute a linear discriminant function to separate the two classes, and obtain a classification rate of 95.7%



A realistic pattern recognition system (4)

- **Cost Versus Classification rate**
  - Our linear classifier was designed to minimize the overall misclassification rate
  - Is this the best objective function for our fish processing plant?
    - The **cost** of misclassifying salmon as sea bass is that the end customer will occasionally find a tasty piece of salmon when he purchases sea bass
    - The **cost** of misclassifying sea bass as salmon is an end customer upset when he finds a piece of sea bass purchased at the price of salmon
  - Intuitively, we could adjust the decision boundary to minimize this cost function

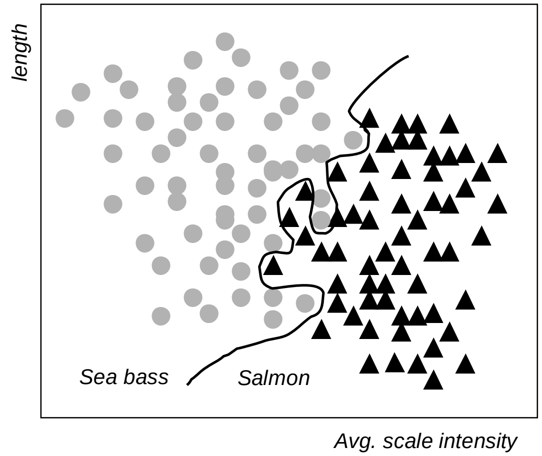




A realistic pattern recognition system (5)

■ The issue of generalization

- The recognition rate of our linear classifier (95.7%) met the design specs, but we still think we can improve the performance of the system
  - We then design an artificial neural network with five hidden layers, a combination of logistic and hyperbolic tangent activation functions, train it with the Levenberg-Marquardt algorithm and obtain an impressive classification rate of 99.9975% with the following decision boundary

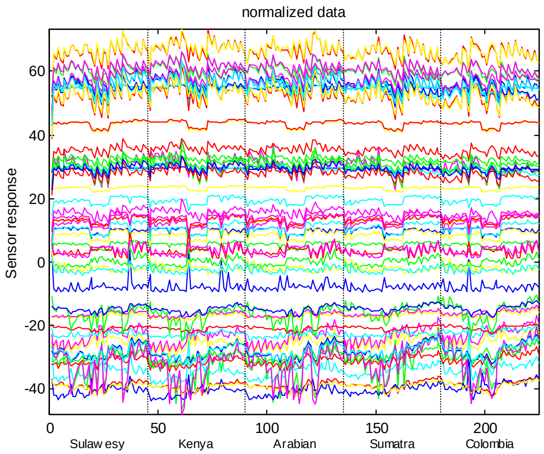


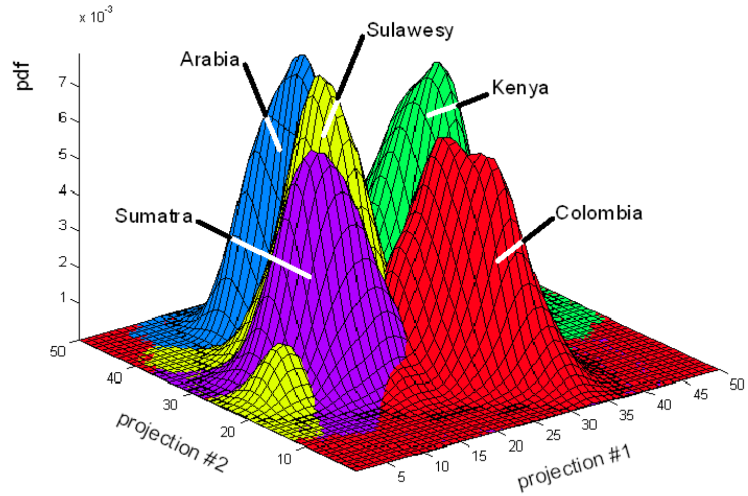
- Satisfied with our classifier, we integrate the system and deploy it to the fish processing plant
  - After a few days, the plant manager calls to complain that the system is misclassifying an average of 25% of the fish
  - What went wrong?

Pattern recognition approaches

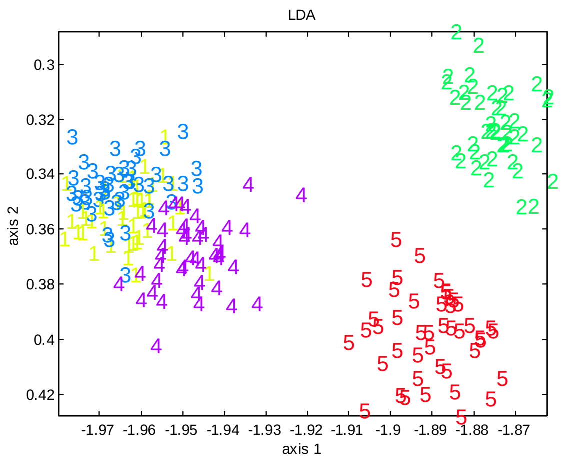
- **Statistical (StatPR)**
  - Patterns classified based on an underlying statistical model of the features
    - The statistical model is defined by a family of class-conditional probability density functions $Pr(x|c_i)$ (Probability of feature vector $\boldsymbol{x}$ given class $\boldsymbol{c_i}$)
- **Syntactic (SyntPR)**
  - Patterns classified based on measures of structural similarity
    - Structure is represented by means of formal grammars or relational descriptions (graphs)
  - SyntPR is used not only for classification, but also for **description**
    - Typically, SyntPR approaches formulate hierarchical descriptions of complex patterns built up from simpler sub patterns.
- **Neural (NeurPR)**
  - Classification is based on the response of a network of processing units (neurons) to an input stimuli (pattern)
    - The response of the network is determined by the connectivity and strength of the synaptic weights
  - NeurPR is a trainable, non-algorithmic, black-box strategy
  - NeurPR is very attractive since
    - it requires minimum a priori knowledge
    - with enough layers and neurons, an ANN can create **any** complex decision region

Statistical pattern recognition: an example



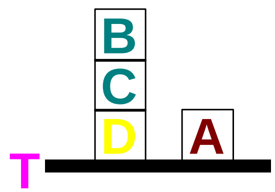
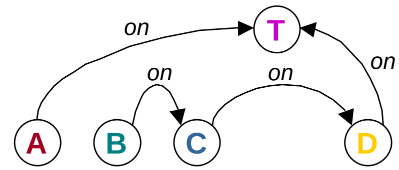


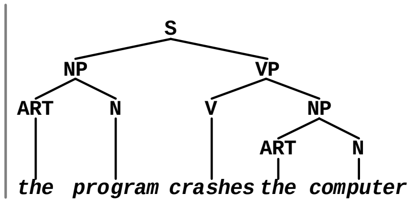
Dimensionality reduction



Density estimation

Structural pattern recognition: an example

Example	Model	Description
	Directed Graph	

`"the program crashes the computer"`	Grammar	`P = {`
		`SENTENCE → NP + VP`
		`NOUN PHRASE → ART + N`
		`VERB PHRASE → V + NP`
		`NOUN → computer|program`
		`VERB → crashes`
		`ARTICLE → the`
		`}`
		

Neural pattern recognition: an example