L1: Course introduction

Introduction

- Course organization
- Grading policy
- Outline

What is pattern recognition?

- Definitions from the literature
- Related fields and applications

Components of a pattern recognition system

- Pattern recognition problems
- Features and patterns
- The pattern recognition design cycle

Pattern Recognition approaches

- Statistical
- Neural
- Structural

Course organization

Instructor

- Ricardo Gutierrez-Osuna
- Office: 520A HRRB
- Tel: (979) 845-2942
- E-mail: rgutier@cse.tamu.edu
- URL: http://faculty.cse.tamu.edu/rgutier

Grading

- Homework
  - 3 assignments, every 3 weeks
- Tests
  - 1 midterm, 1 final (comprehensive)
- Term project
  - Open-ended
  - Public presentation

	Weight (%)
Homework	40
Project	30
Midterm	15
Final Exam	15

What is pattern recognition?

Definitions from the literature

- "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*
- "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 $\mathbf{x}$", *–Schürmann*
- Pattern Recognition is concerned with answering the question "***What is this?***" *–Morse*

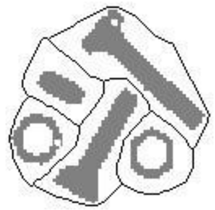
Examples of pattern recognition problems

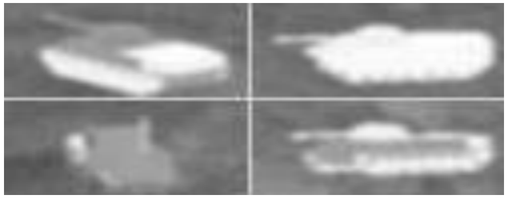
Machine vision
- Visual inspection, ATR
- Imaging device detects ground target
- Classification into “friend” or “foe”

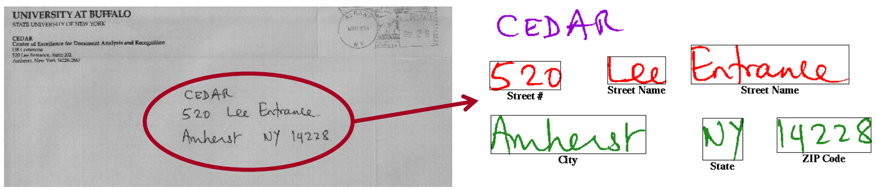
Character recognition
- Automated mail sorting, processing bank checks
- Scanner captures an image of the text
- Image is converted into constituent characters

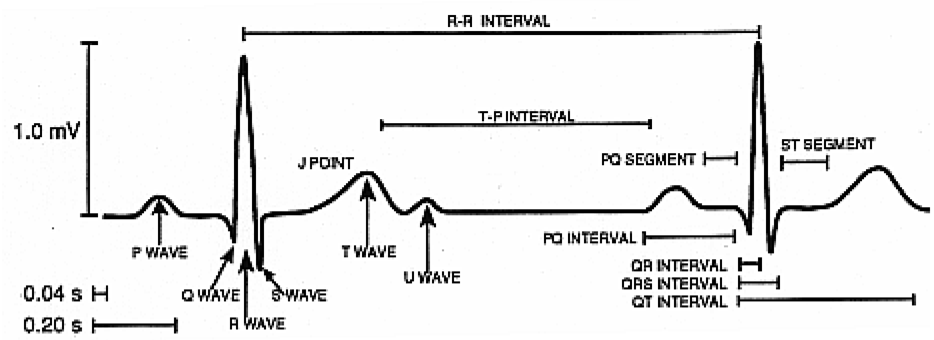
Computer aided diagnosis
- Medical imaging, EEG, ECG signal analysis
- Designed to assist (not replace) physicians
- Example: X-ray mammography
  - 10-30% false negatives in x-ray mammograms
  - 2/3 of these could be prevented with proper analysis

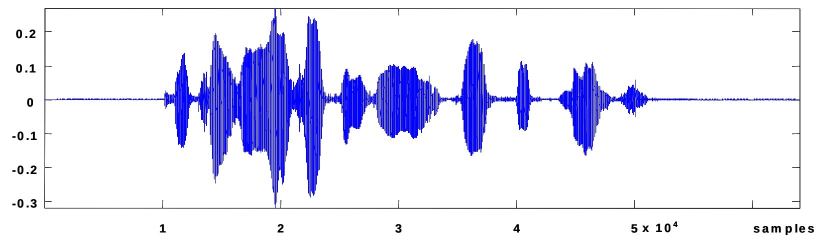
Speech recognition
- Human Computer Interaction, Universal Access
- Microphone records acoustic signal
- Speech signal is classified into phonemes and/or words











Related fields and application areas for PR

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
- Biological cybernetics
- Computational neuroscience

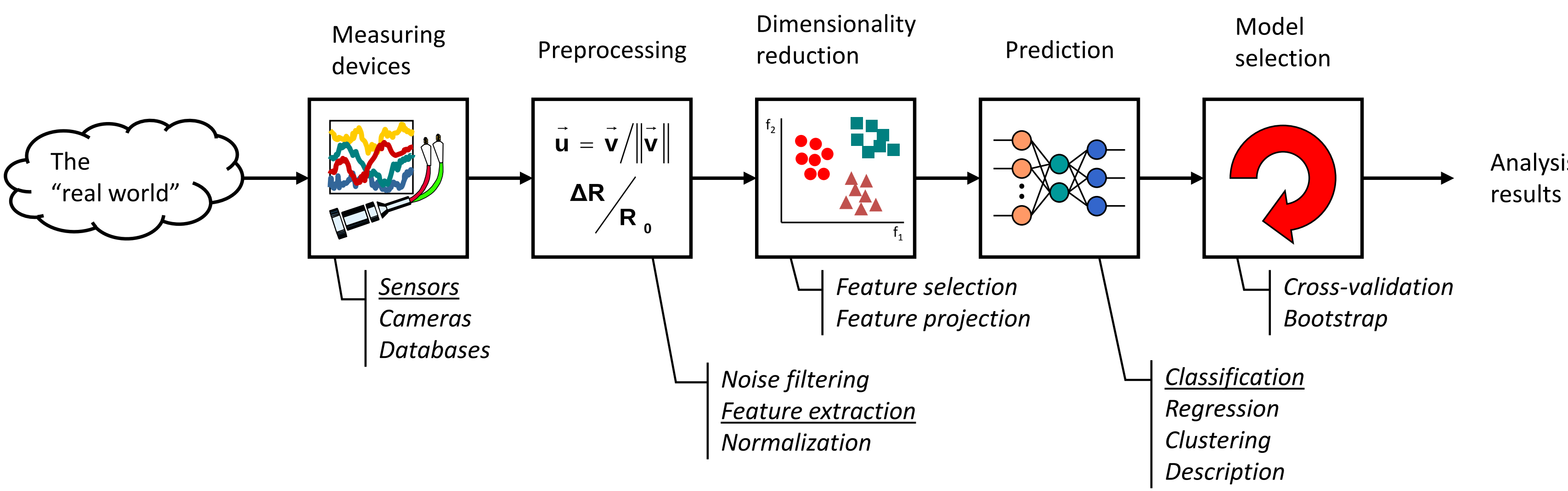
Applications

- Image processing
- Computer vision
- Speech recognition
- Multimodal interfaces
- Automated target recognition
- Optical character recognition
- Seismic analysis
- Man and machine diagnostics
- Fingerprint identification
- Industrial inspection
- Financial forecast
- Medical diagnosis
- ECG signal analysis

Components of a pattern recognition system

A basic pattern classification system contains

- A sensor
- A preprocessing mechanism
- A feature extraction mechanism (manual or automated)
- A classification algorithm
- A set of examples (training set) already classified or described



Types of prediction problems

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

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

Features and patterns

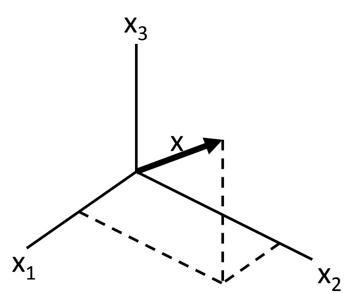
Feature

- 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 features is a d-dim 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; the result is a scatter plot

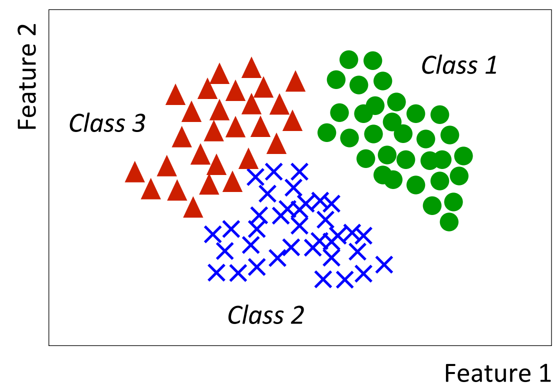
Feature vector

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

Feature space (3D)



Scatter plot (2D)

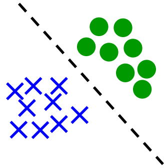


Pattern

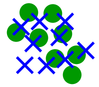
- Pattern is a composite of traits or features characteristic of an individual
- In classification tasks, a pattern is a pair of variables $\{x, \omega\}$ where
  - x is a collection of observations or features (feature vector)
  - ω is the concept behind the observation (label)

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

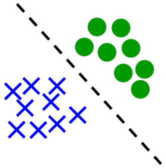


"Good" features

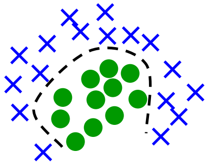


"Bad" features

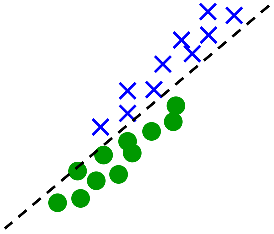
More feature properties



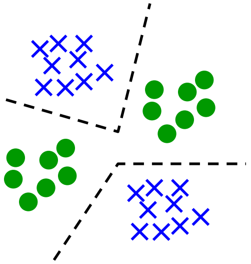
Linear separability



Non-linear separability



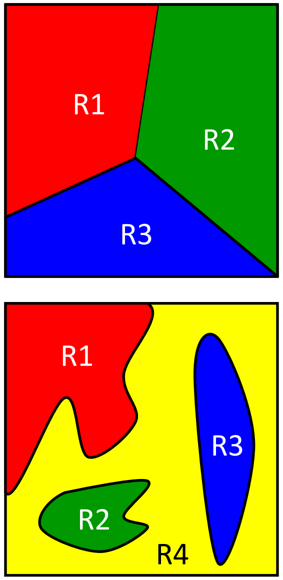
Highly correlated features



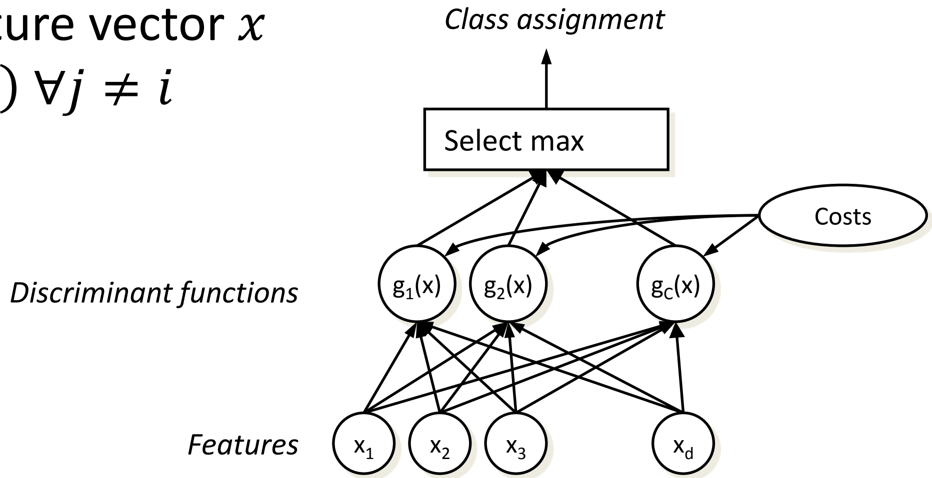
Multi-modal

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 $g_i(x) > g_j(x)\ \forall j \neq i$



Pattern recognition approaches

Statistical

- 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 $p(x|\omega_i)$ (Probability of feature vector x given class ω_i)

Neural

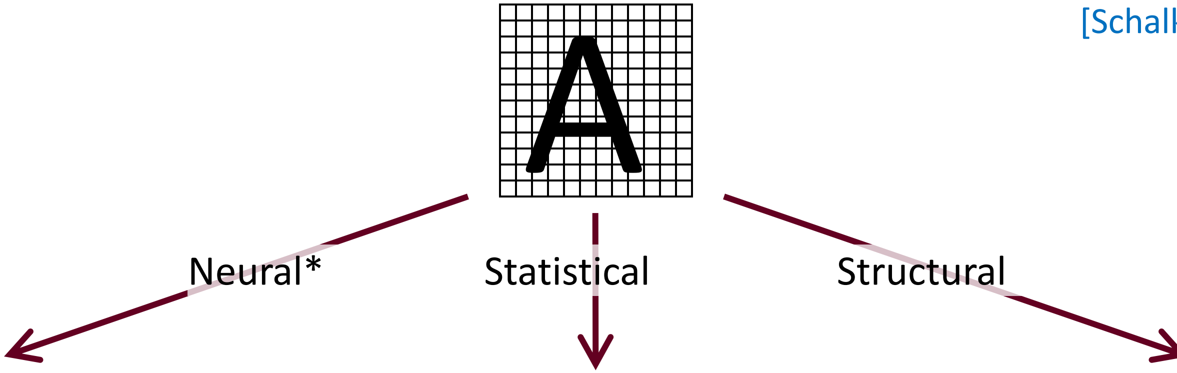
- Classification is based on the response of a network of processing units (neurons) to an input stimuli (pattern)
  - "Knowledge" is stored in the **connectivity and strength of the synaptic weights**
- Trainable, non-algorithmic, black-box strategy
- Very attractive since
  - it requires minimum a priori knowledge
  - with enough layers and neurons, ANNs can create **any** complex decision region

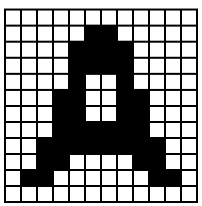
Syntactic

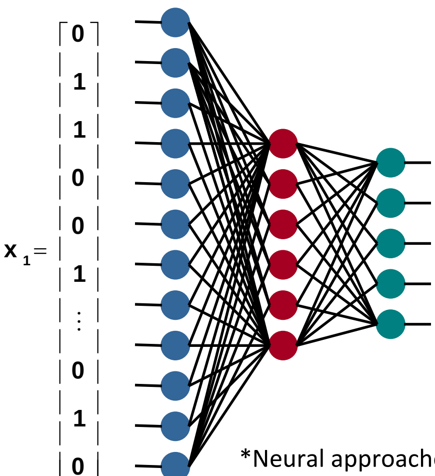
- Patterns classified based on measures of structural similarity
  - "Knowledge" is represented by means of **formal grammars or relational descriptions** (graphs)
- Used not only for classification, but also for description
  - Typically, syntactic approaches formulate hierarchical descriptions of complex patterns built up from simpler sub patterns

Example: neural, statistical and structural OCR

[Schalkoff, 1992]



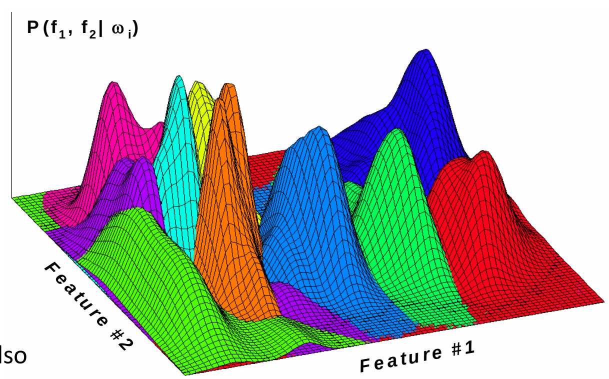


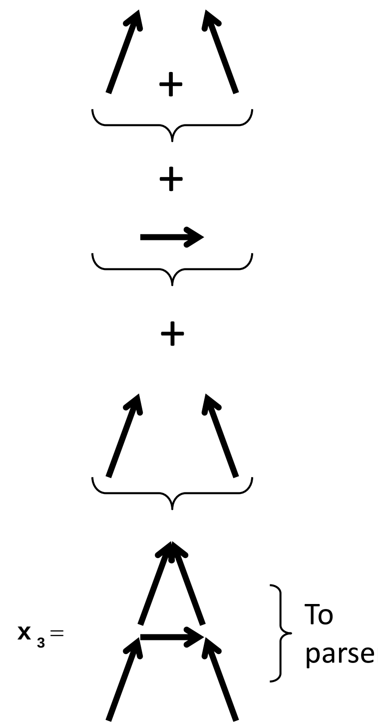


*Neural approaches may also employ feature extraction

Feature extraction:
- # intersections
- # right oblique lines
- # left oblique lines
- # horizontal lines
- # "holes"

$$x_2 = [3 \quad 2 \quad 1 \quad 2 \quad 1]^T \xrightarrow{\text{Probabilistic model}} p(x_2 | \text{"A"})$$





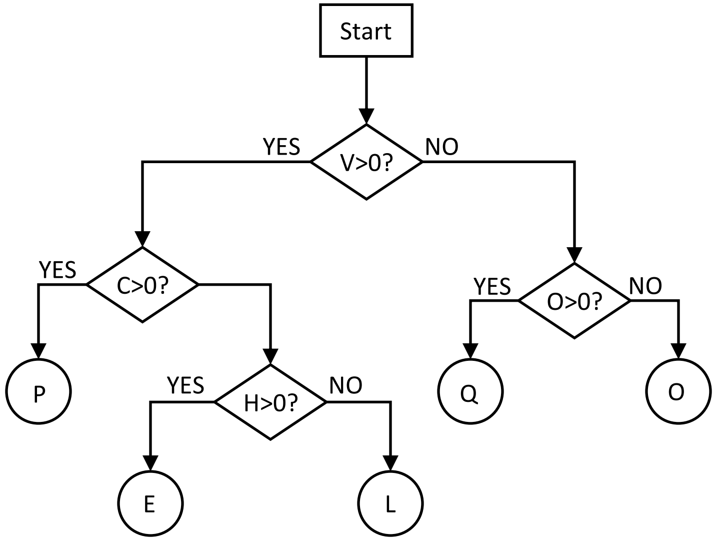
A simple pattern recognition problem

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



The pattern recognition design cycle

Data collection
- Probably the most time-intensive component of a PR project
- How many examples are enough?

Feature choice
- Critical to the success of the PR problem
  - "Garbage in, garbage out"
- Requires basic prior knowledge

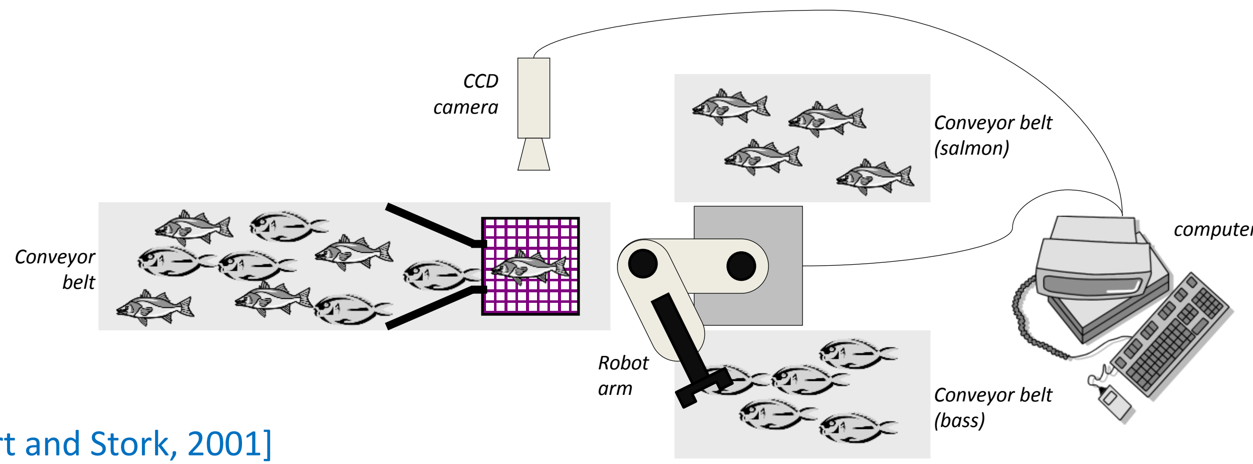
Model choice
- Statistical, neural and structural approaches
- Parameter settings

Training
- Given a feature set and a "blank" model, adapt the model to explain the data
- Supervised, unsupervised and reinforcement learning

Evaluation
- How well does the trained model do?
- Overfitting vs. generalization

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



[Duda, Hart and Stork, 2001]

Sensor

– The vision system captures an image as a new fish enters the sorting area

Preprocessing

– Image processing algorithms, e.g., 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

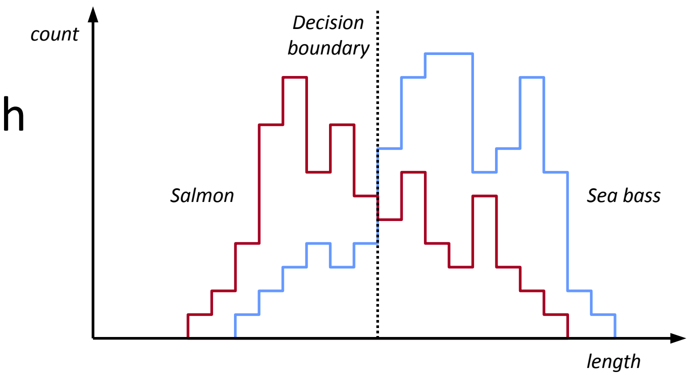
– Collect a set of examples from both species

– Compute the distribution of lengths for both classes

– Determine a decision boundary (threshold) that minimizes the classification error

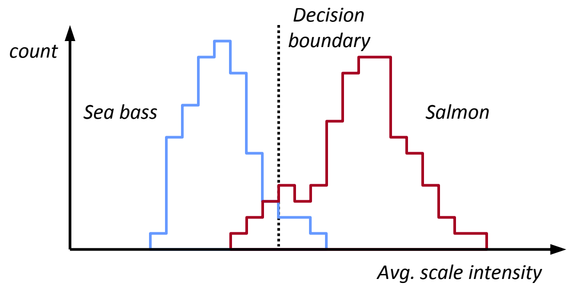
– We estimate the classifier's probability of error and obtain a discouraging result of 40%

– **What do we do now?**

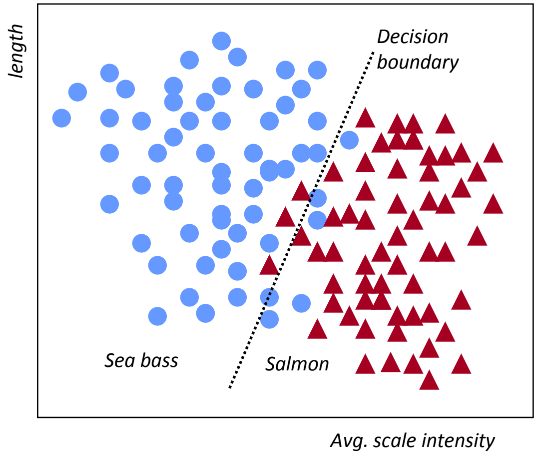


Improving the performance of our PR system

- 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

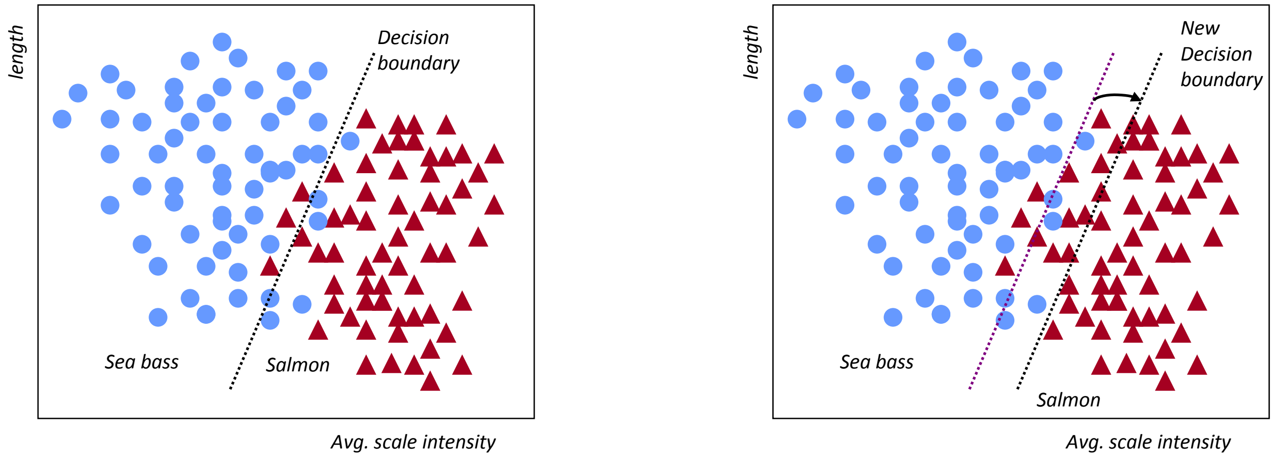


- We combine “length” and “average intensity of the scales” to improve class separability
- We compute a linear discriminant function to separate the two classes, and obtain a classification rate of 95.7%



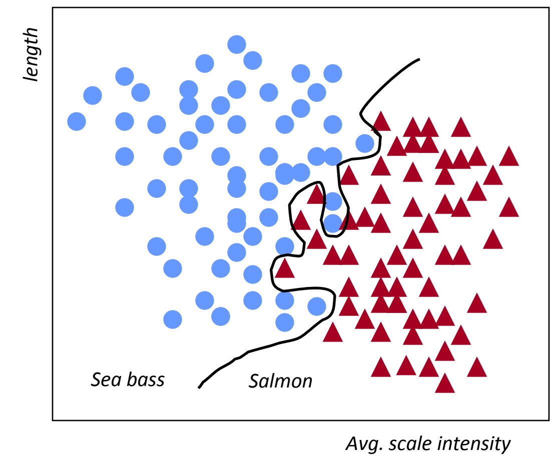
Cost vs. classification rate

- 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



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 ANN 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?