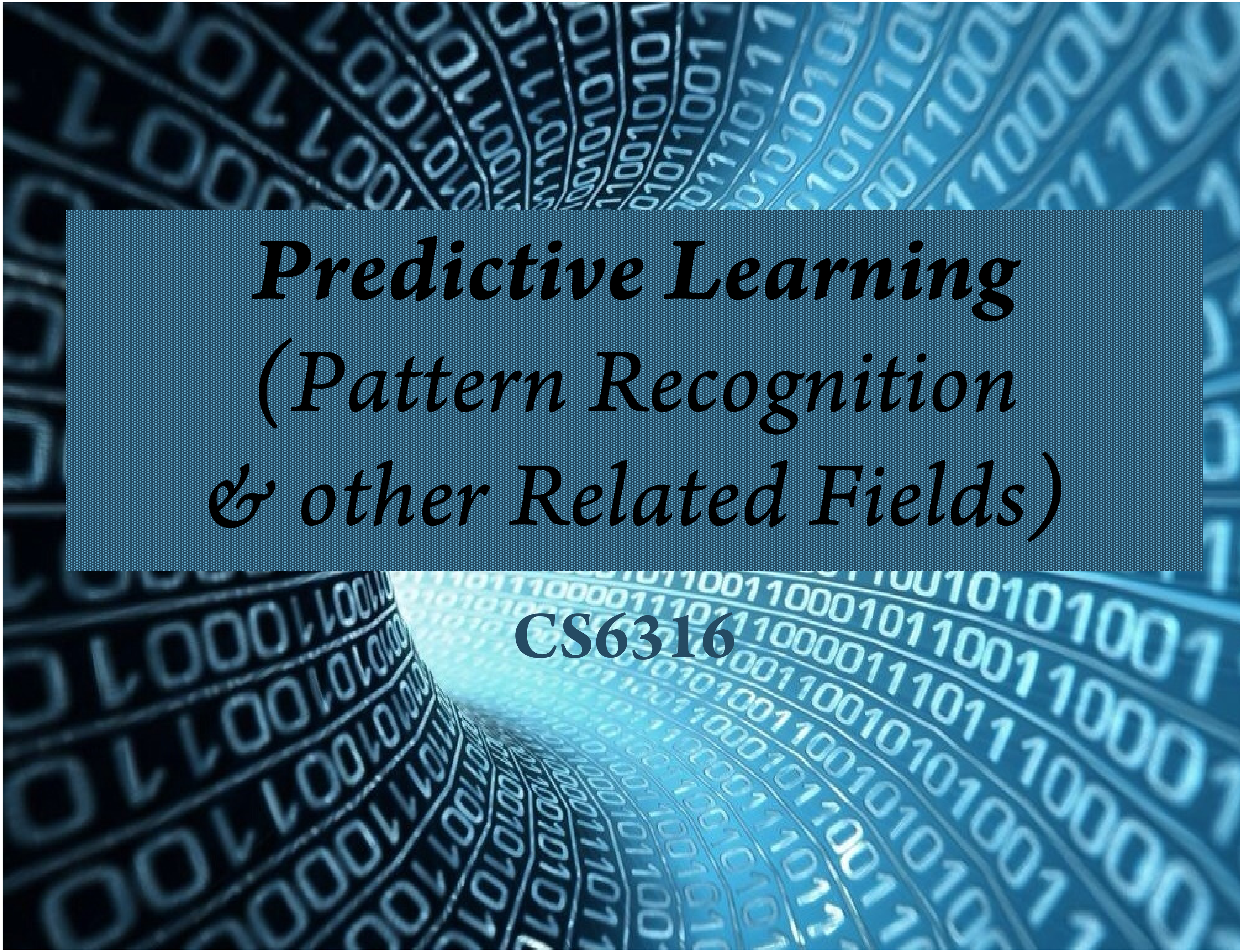

Machine Learning
CS6316



Predictive Learning (Pattern Recognition & other Related Fields)

CS6316

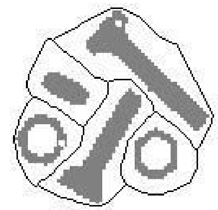
"In a learning system, categorical outputs occur quite often, and represent a class of learning problems known as ***PATTERN RECOGNITION*** *or* ***CLASSIFICATION*****."**

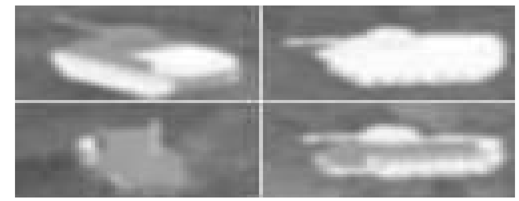
Related Fields & Applications

Pattern Recognition

- "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 of 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

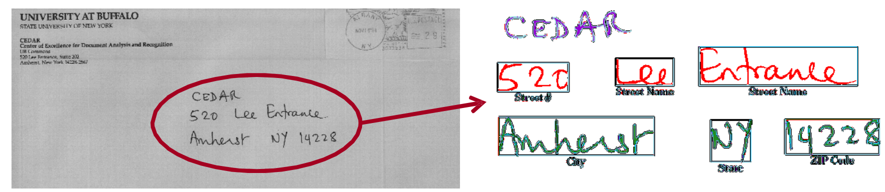




Machine vision

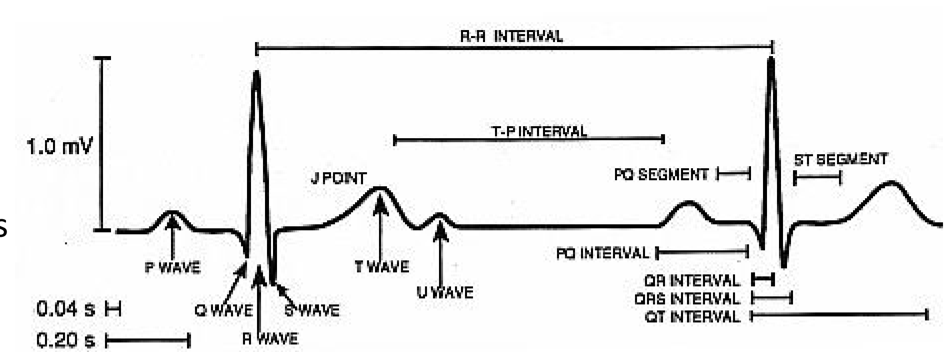
- Visual inspection, ATR
- Imaging device detects ground target
- Classification into "friend" or "foe"

Character recognition



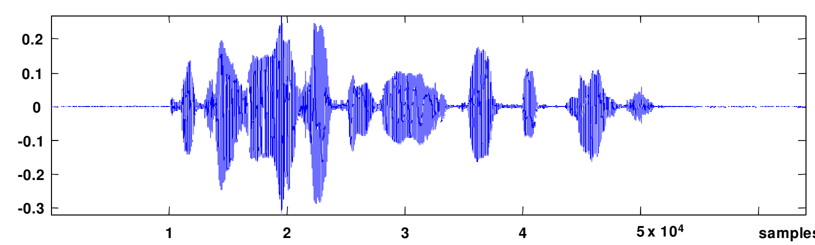
- Automated mail sorting, processing bank checks
- Scanner captures an image of the text
- Image is converted into constituent characters

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 & Application Areas

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

- Imagine 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

Types of Prediction Problems

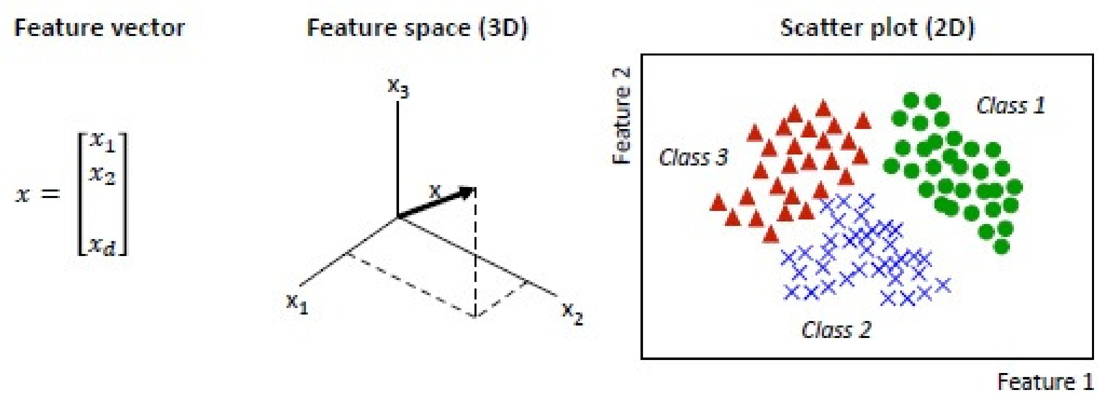
- **Classification**
  - The problem of assigning an object/item to a class
  - The output of the system is a label (typically an integer)
    - E.g. classifying a product as "good" or "bad" in a quality control test
- **Regression**
  - A generalization of a classification task
  - The output of the 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

- **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 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**
  - Any distinctive aspect, quality or characteristic
    - 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

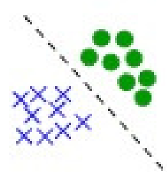


Features and Patterns

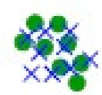
- **Feature**
  - Pattern is a composite of traits, or features characteristic of an individual
  - In classification tasks, a pattern is a pair of variables $\{x, \omega\}$ where
    - x is a collection of observations or features (feature vector)
    - ω is the concept behind the observation (label)

What makes a “good” feature vector?

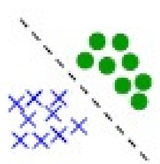
- The quality of a feature vector is related to its ability to discriminate examples from different classes
  - Examples from the same class should have similar feature values
  - Examples from different classes have different feature values



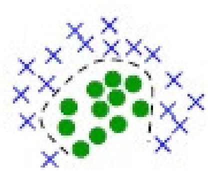
“Good” features



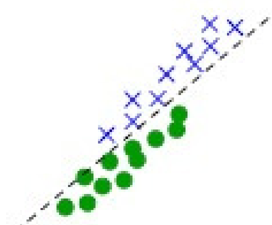
“Bad” features



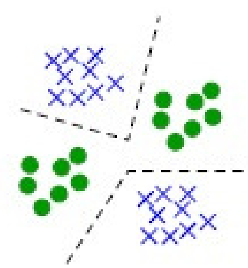
Linear separability



Non-linear separability



Highly correlated features

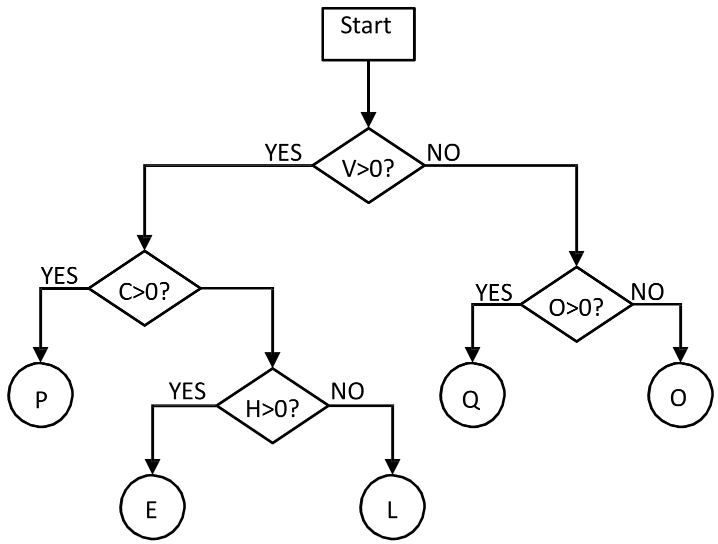


Multi-modal

Simple Case Study: Letter Recognition

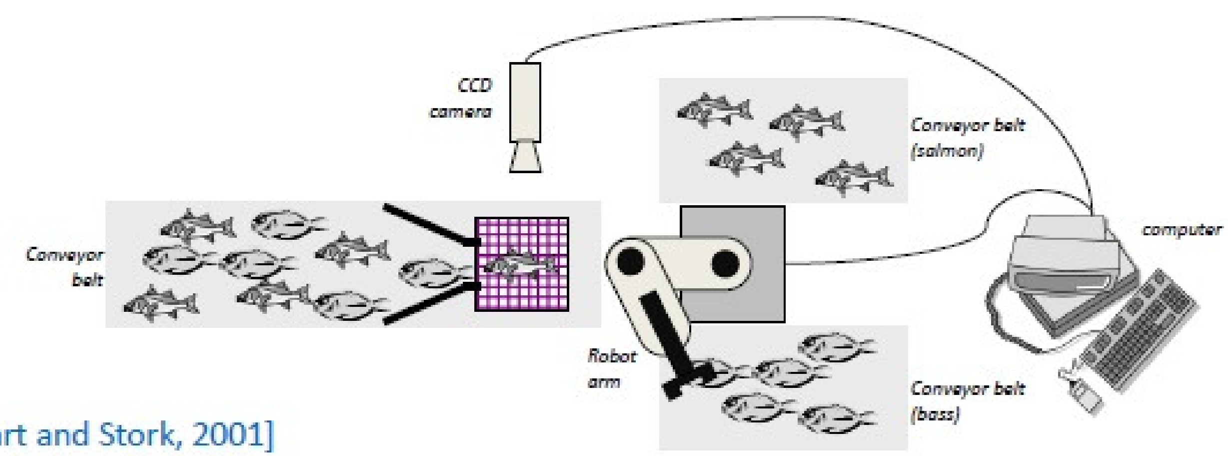
- Consider the problem of recognizing the letters L, P, O, E, Q
  - Determine a sufficient set of features
  - Design a tree-structure classifier

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



Another Case Study: Salmon vs Sea Bass

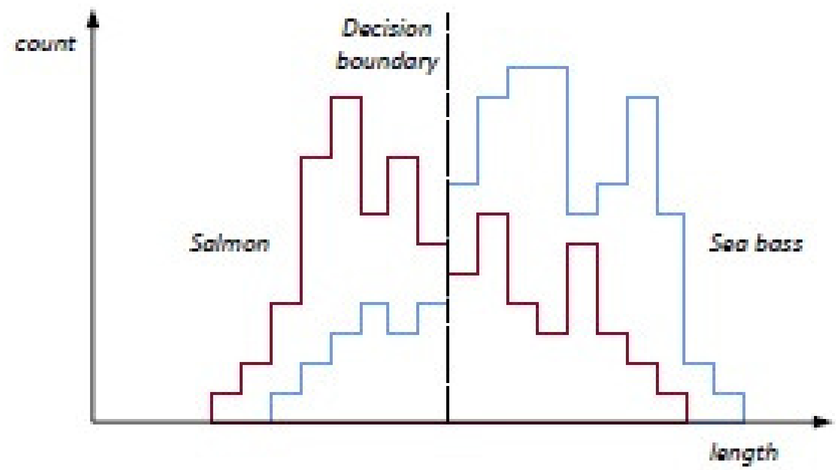
- A fish processing plan wants to automate the process of sorting incoming fish according to species (salmon or sea bass)
- The automation system consists of:
  - a conveyor belt for incoming products
  - two conveyor belts for sorted products
  - a pick-and-place robotic arm
  - a vision system with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



[Duda, Hart and Stork, 2001]

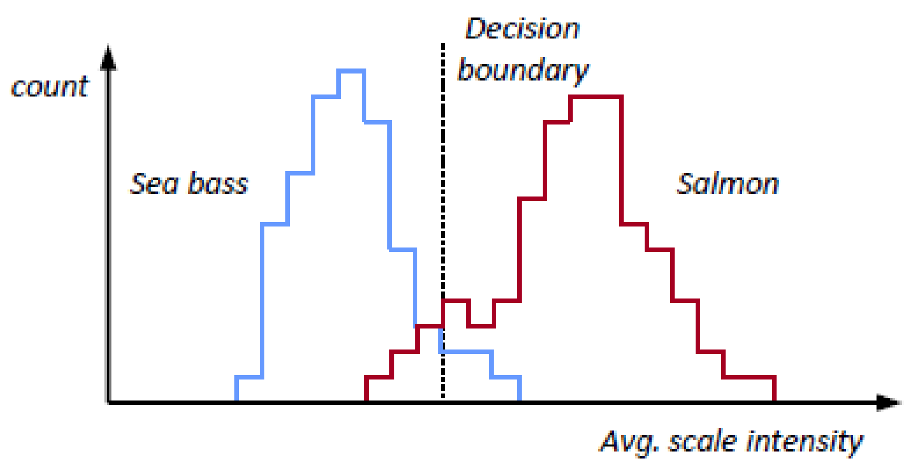
- **Sensor**
  - The vision system captures an image as a new fish enters the sorting area
- **Preprocessing**
  - Image processing algorithms e.g. adjustments for average intensity levels, segmentation to separate fish from background
- **Feature extraction**
  - Suppose we know that, on the average, 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?



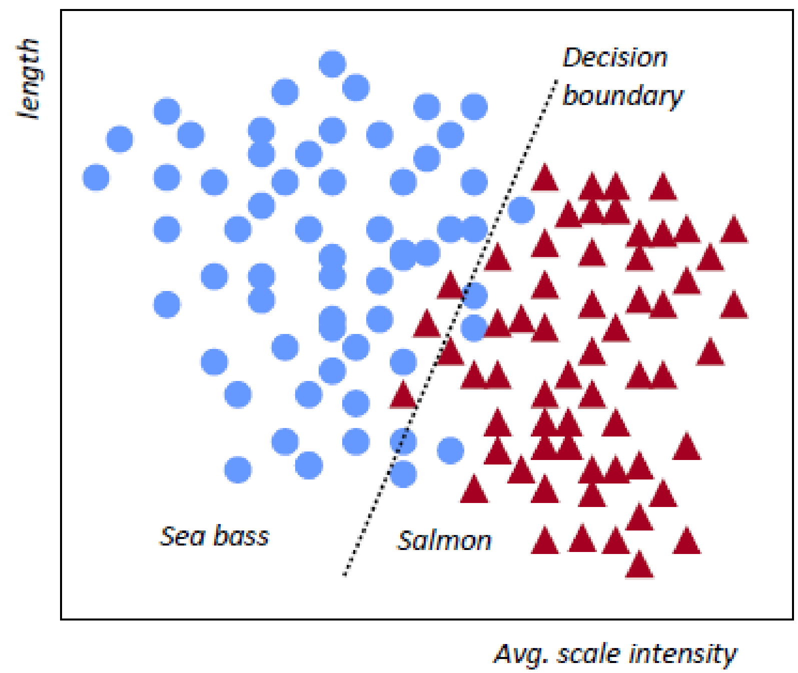
Improving the performance…

- 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



Improving the performance…

- 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%

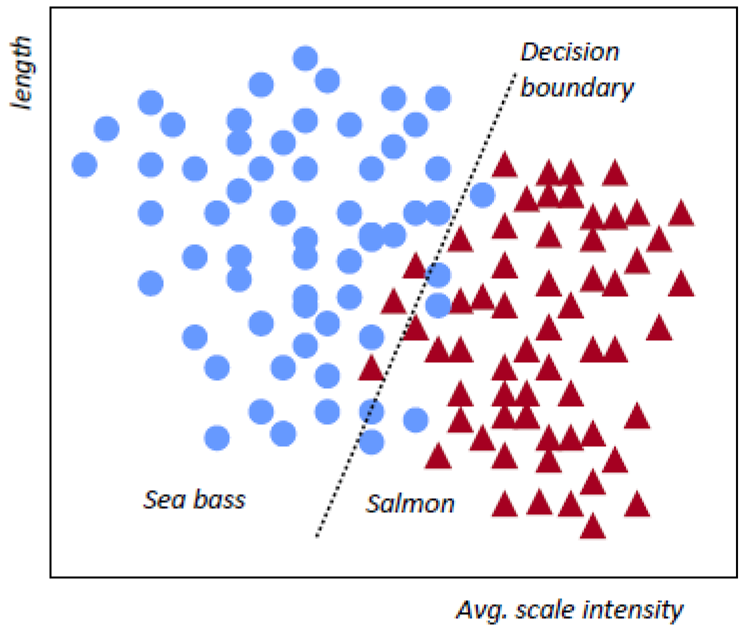


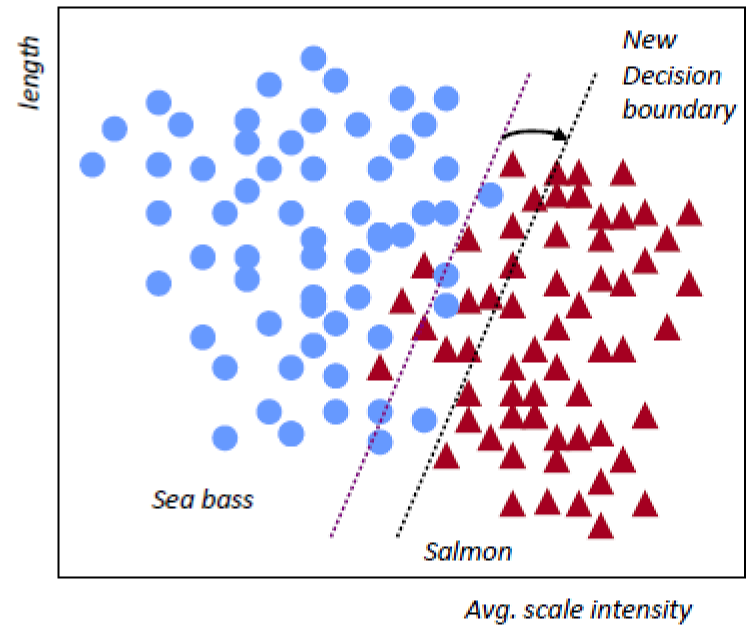
Cost vs. Classification Rate

- The linear classifier (see previous slide) was designed to minimize the overall misclassification rate
- Is this the best objective function for the 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*

Cost vs. Classification Rate

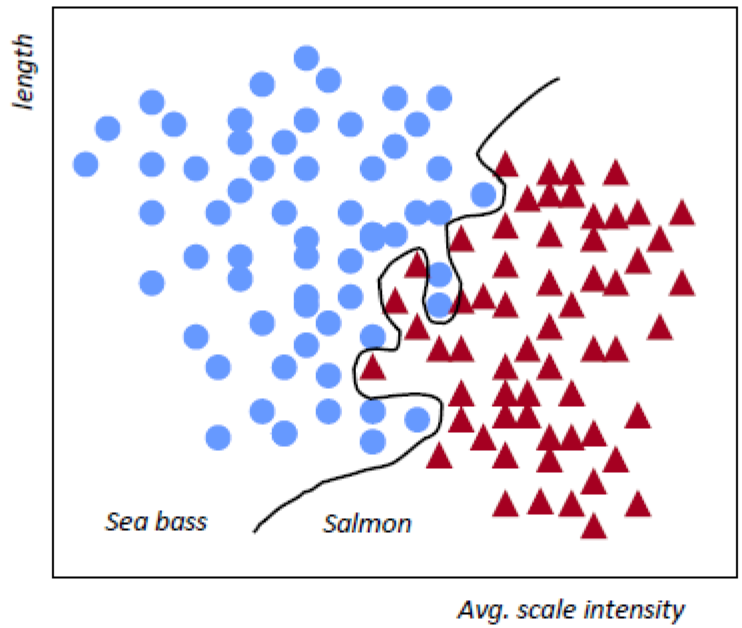
- Intuitively, we could adjust the decision boundary to minimize this cost function





The Issue of Generalization

- The recognition rate of the 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



- Now, satisfied with our classifier, we integrate the system and deploy it to the fish processing plant

The Issue of Generalization

- 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??