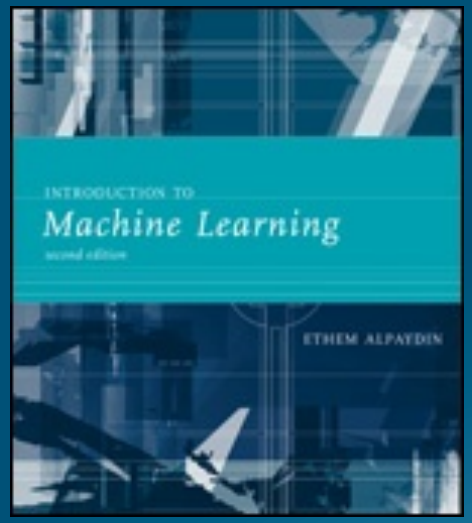


Lecture Slides for

INTRODUCTION TO

Machine Learning

2nd Edition

ETHEM ALPAYDIN

alpaydin@boun.edu.tr
http://www.cmpe.boun.edu.tr/~ethem/i2ml2e

In preparation of these slides, I have benefited from slides prepared by:

E. Alpaydin (Intro. to Machine Learning),

D. Bouchaffra and V. Murino (Pattern Classification and Scene Analysis),

R. Gutierrez-Osuna (Texas A&M)

A. Moore (CMU)

CHAPTER 1:

Introduction

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About"Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Blink" also bought "Outliers" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.

Data Mining

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- ...

What is Machine Learning?

- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

Learning Associations

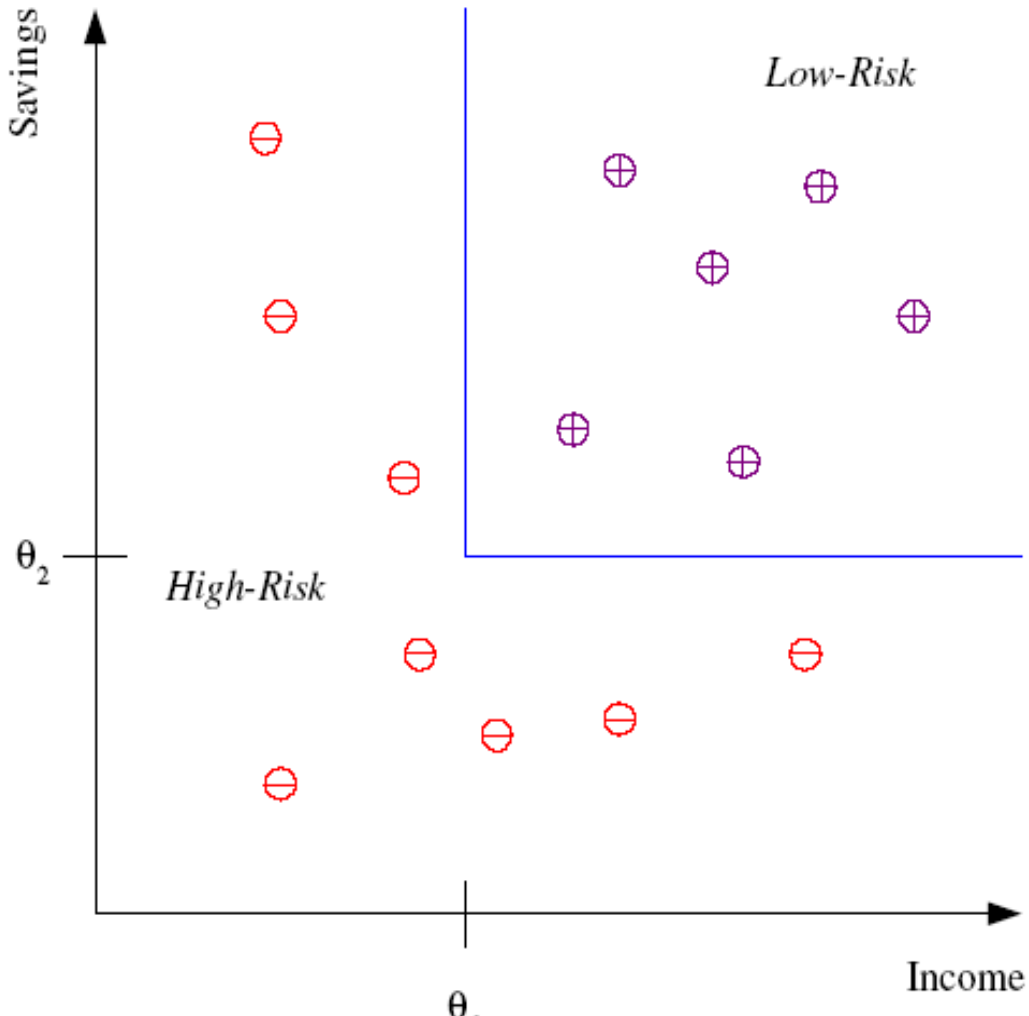
- Basket analysis:

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  Example: $P(\text{chips} \mid \text{beer}) = 0.7$

Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

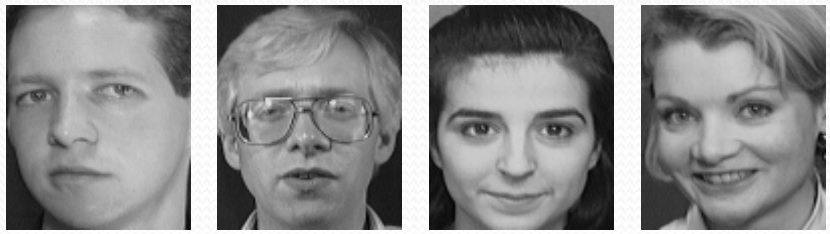
Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- ...

Face Recognition

Training examples of a person

Test images



ORL dataset,
AT&T Laboratories, Cambridge UK

Regression

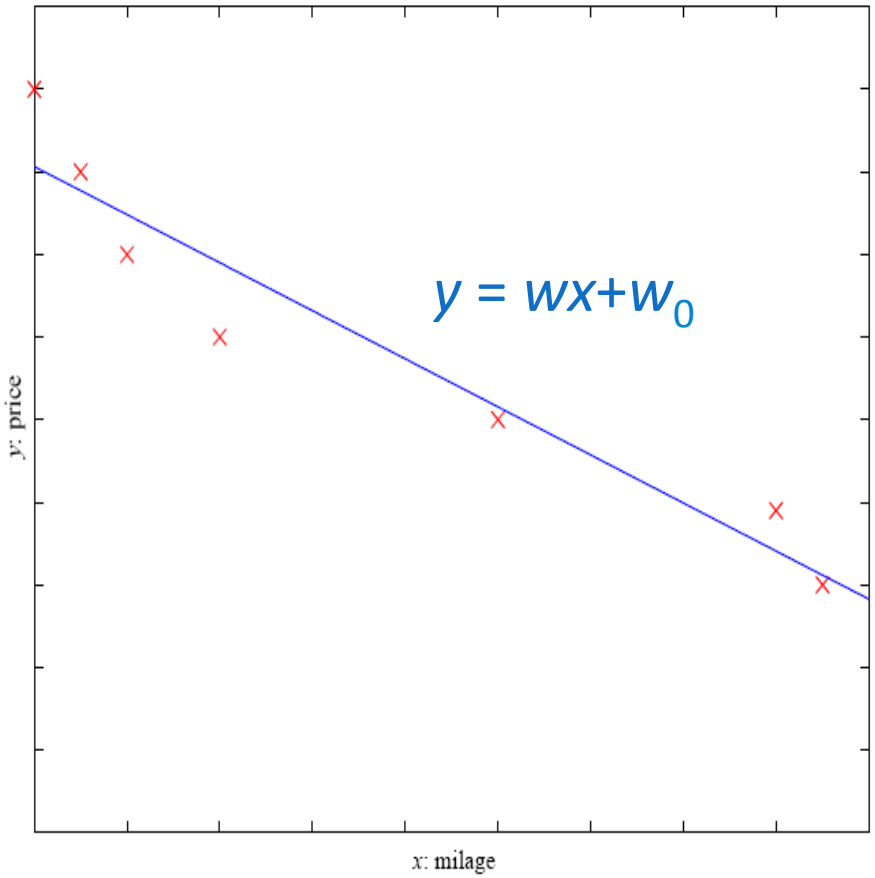
- Example: Price of a used car
- x : car attributes

  y : price

  $$y = g(x \mid \theta)$$

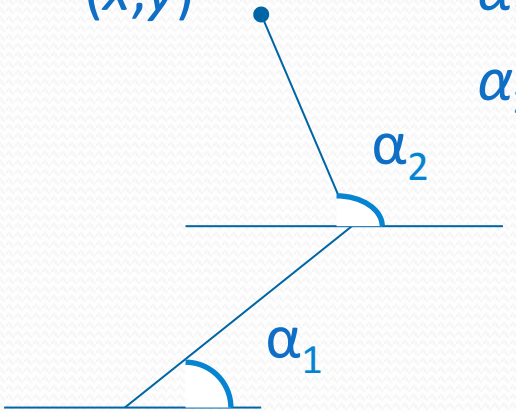
  $g(\ )$ model,

  θ parameters



Regression Applications

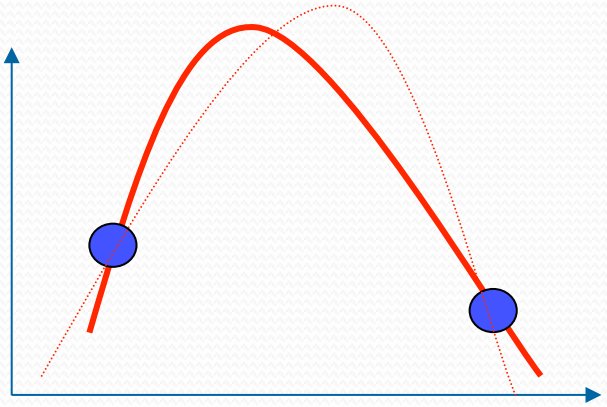
- Navigating a car: Angle of the steering
- Kinematics of a robot arm



$\alpha_1 = g_1(x,y)$

$\alpha_2 = g_2(x,y)$

- Response surface design



Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

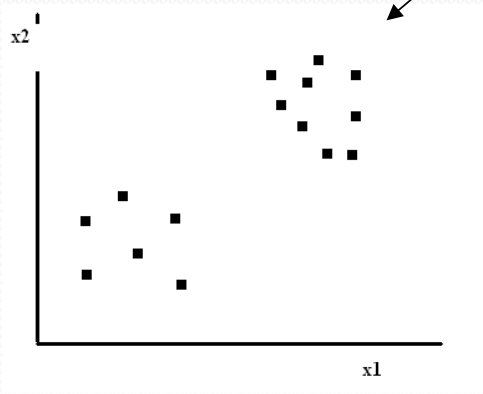
Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

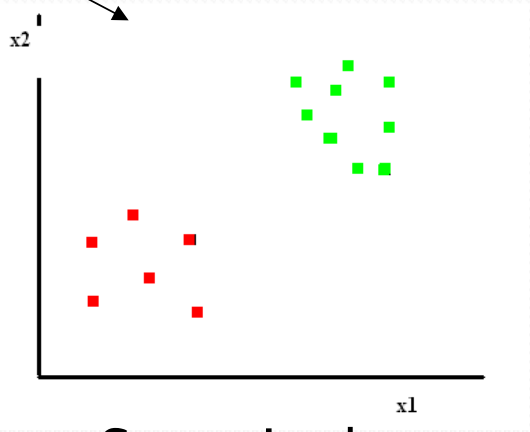
Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

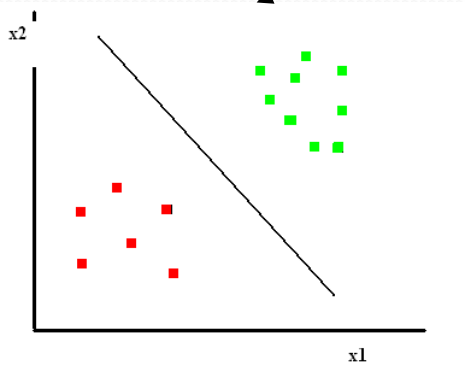
ML Aplications



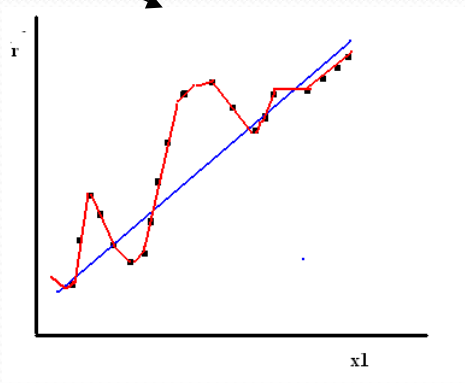
Unsupervised



Supervised



Classification

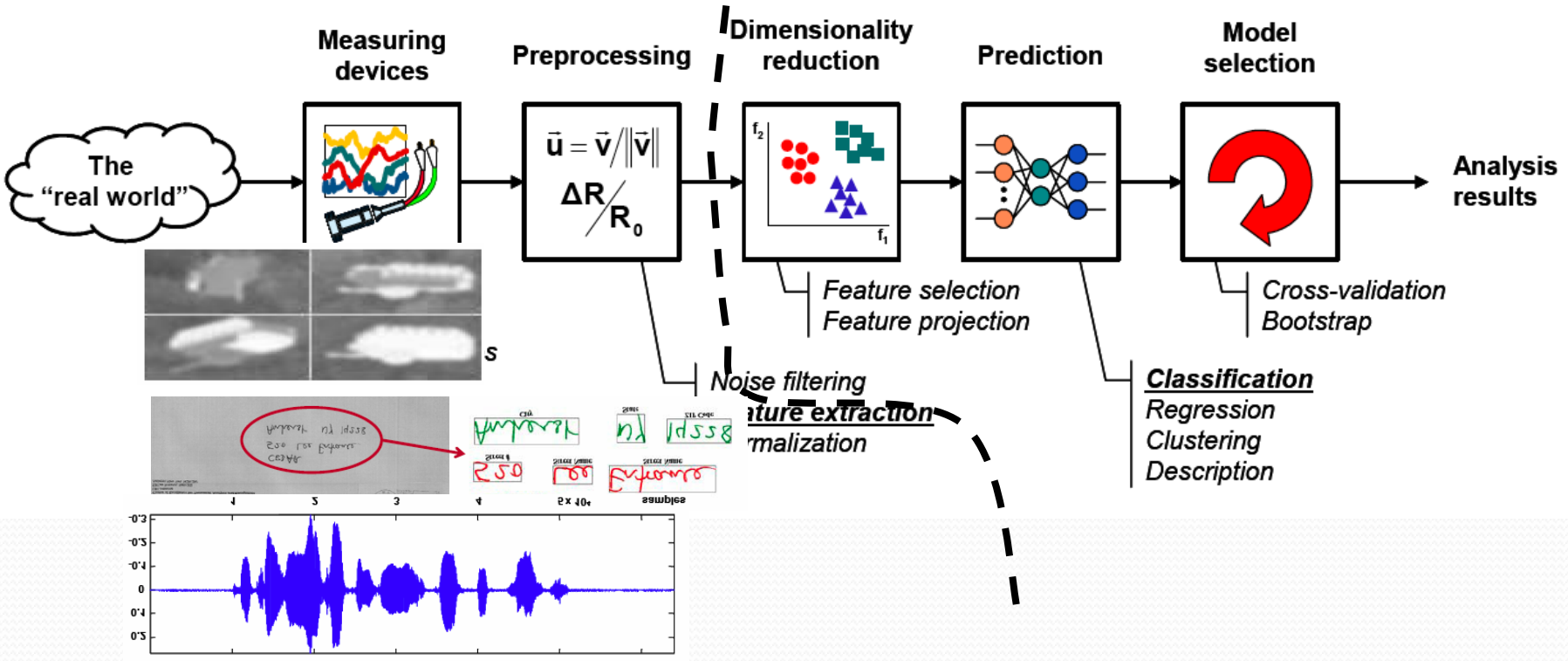


Regression

Components of a PR System

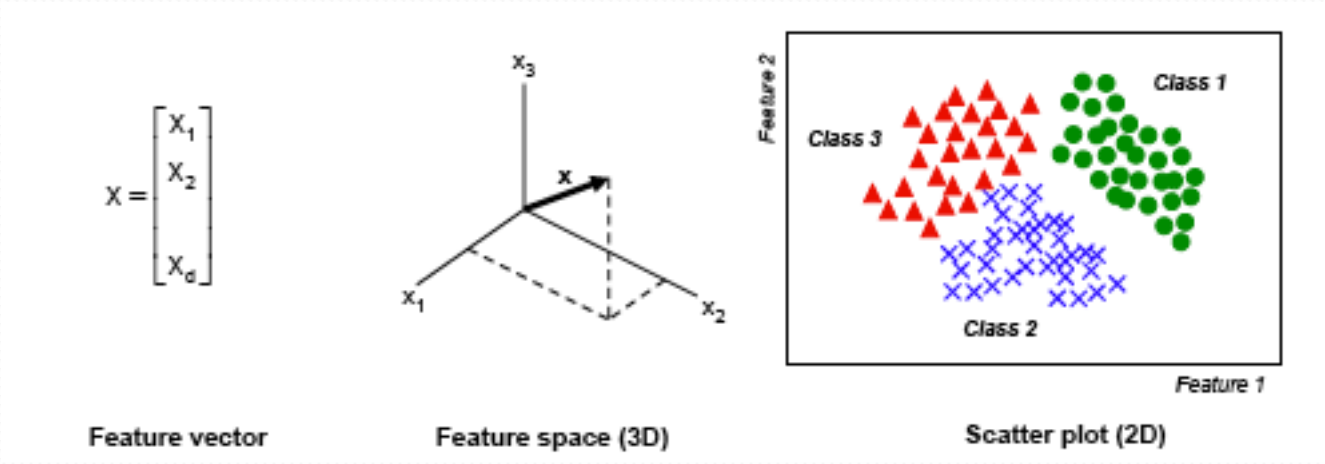
A basic pattern classification system contains

- *A sensor, preprocessing and feature extraction mechanism (manual or automated)*
- Dimensionality reduction step
- A classification (regression, clustering, description) algorithm



Features and patterns (1)

- Feature is any distinctive aspect, quality or characteristic
  - Features may be symbolic (i.e., color) or numeric (i.e., height)
  - 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

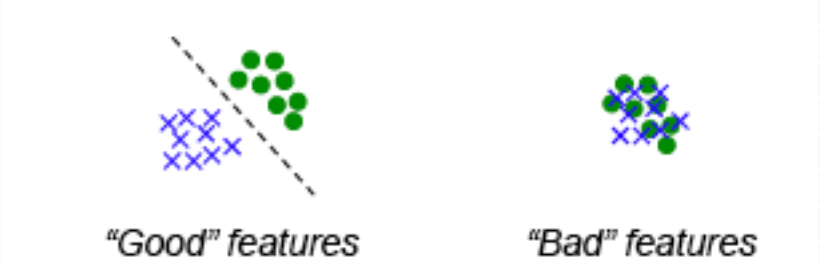


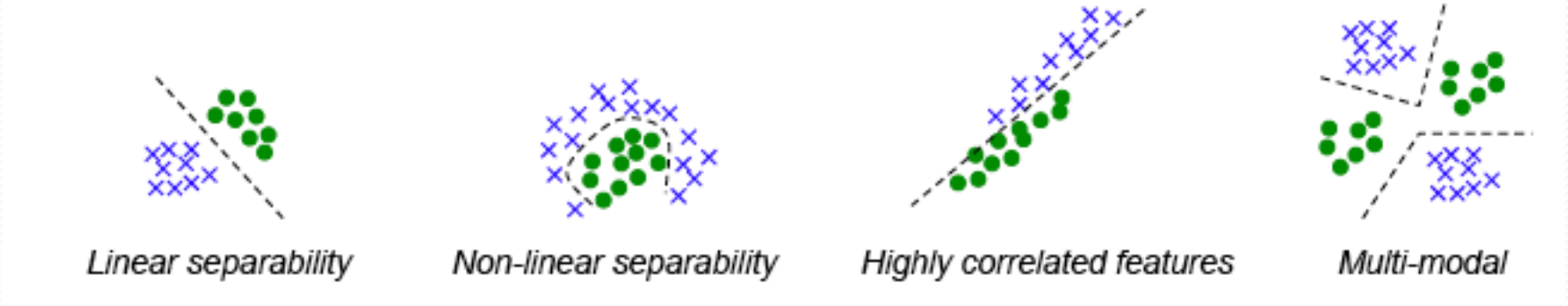
Feature vector | Feature space (3D) | Scatter plot (2D)

- Pattern is a composite of traits or features characteristic of an individual
  - In classification tasks, a pattern is a pair of variables {x,r} where
    - x is a collection of observations or features (feature vector)
    - r is the concept behind the observation (label) (sometimes we will use t instead of r)

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



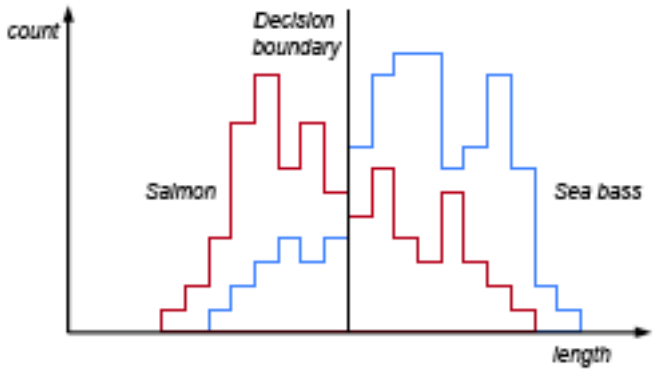


PR design cycle

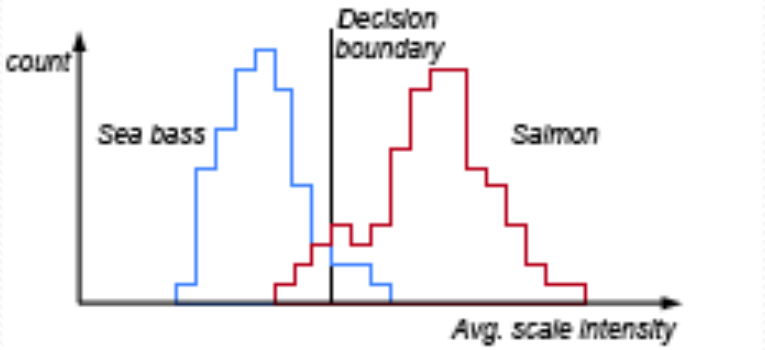
- Data collection
  - Probably the most time-intensive component of a PR project
  - How many examples are enough?
- Feature choice
  - Critical to the success of the PR problem
  - “Garbage in, garbage out”
  - Requires basic prior knowledge
- Model choice
  - Statistical, neural and structural approaches
  - 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

Feature Selection

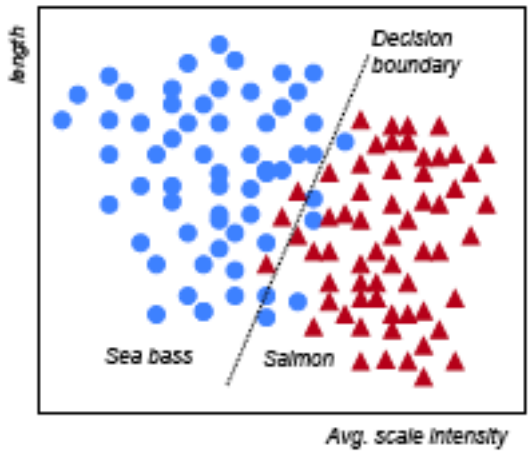
(Salmon vs Sea Bass Recognition Problem)



Length



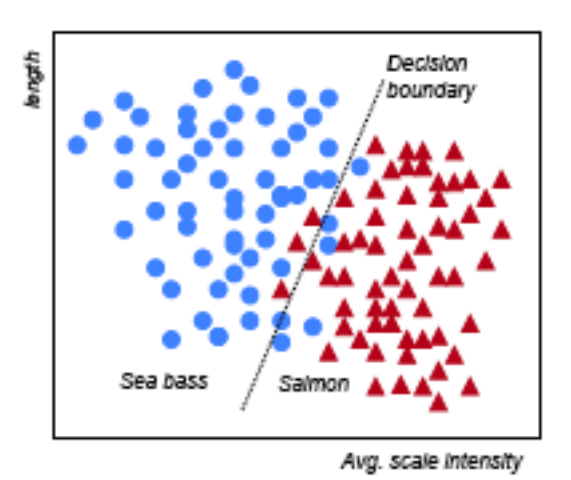
Avg. scale intensity



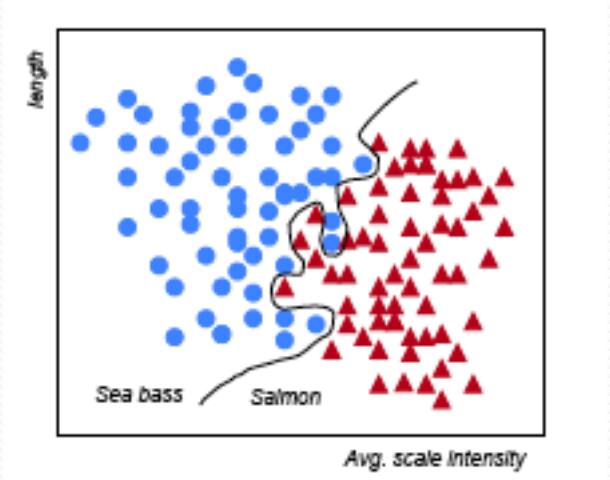
Length and Avg. scale intensity

Model Selection

(Salmon vs Sea Bass Recognition Problem)



Linear Discriminant Function
Performance: 95.7%



Nonlinear (Neural Network) Function
Performance: 99.9%

Which model should we use?
Training vs. Test performance.

Resources: Libraries

- Python-scikit (UI, lib)
- Weka (UI, lib java based)
- R (UI, lib)
- Mlpack (C)
- Spark Mllib (scala, java)
- Matlab-prtools, nntools?
- Knime (interesting UI)

Resources: Datasets

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve/
- Kaggle

Resources: Journals (scholar.google)

- Journal of Machine Learning Research
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

Resources: Conferences (scholar.google)

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- International Conference on Pattern Recognition (ICPR)
- ICMLA, ICDM, ICDE, ...

Resources: Video Lectures

Lectures:
Yaser S. Abu-Mostafa:
http://work.caltech.edu/telecourse.html

Andrew Ng:http://www.academicearth.org/courses/machine-learning
http://see.stanford.edu/see/lecturelist.aspx?coll=348ca38a-3a6d-4052-937d-cb017338d7b1

Tom Mitchell:http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml

Anil Jain:
http://ocw.korea.edu/ocw/college-of-engineering/introduction-to-pattern-recognition

Statistical Learning:
Deep NN….
Latest topics:

http://videolectures.net/Top/Computer_Science/Machine_Learning/