Pattern Recognition

Lecture 1: Introduction

Ahmed Hamdy

Computer Engineering Department

Cairo University

Spring 2017

What is PR?

- ***Pattern recognition*** is a branch of machine learning that focuses on the recognition of patterns and regularities in data.

- Sometimes synonymous with machine learning.

PR applications

Problem Domain	Application	Input Pattern	Pattern Classes
Bioinformatics	Sequence Analysis	DNA/Protein sequence	Known types of genes/ patterns
Data mining	Searching for meaningful patterns	Points in multi-dimensional space	Compact and well-separated clusters
Document classification	Internet search	Text document	Semantic categories (e.g., business, sports, etc.)
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters, words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective / non-defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres (e.g., action, dialogue, etc.)
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory enquiry without operator assistance	Speech waveform	Spoken words

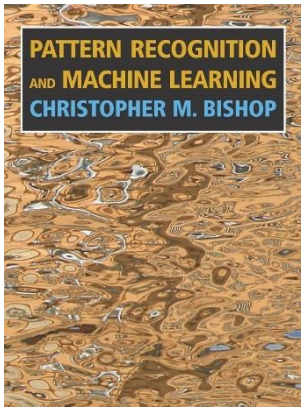
Courtesy of Anil Jain et al.

Topics in PR

- Supervised Learning
  - Regression
  - Classification (Logistic Regression, SVM, KNN, GP,…)
- Regularization and Model Selection
- Unsupervised Learning (K-means clustering)
- Reinforcement Learning

Textbooks

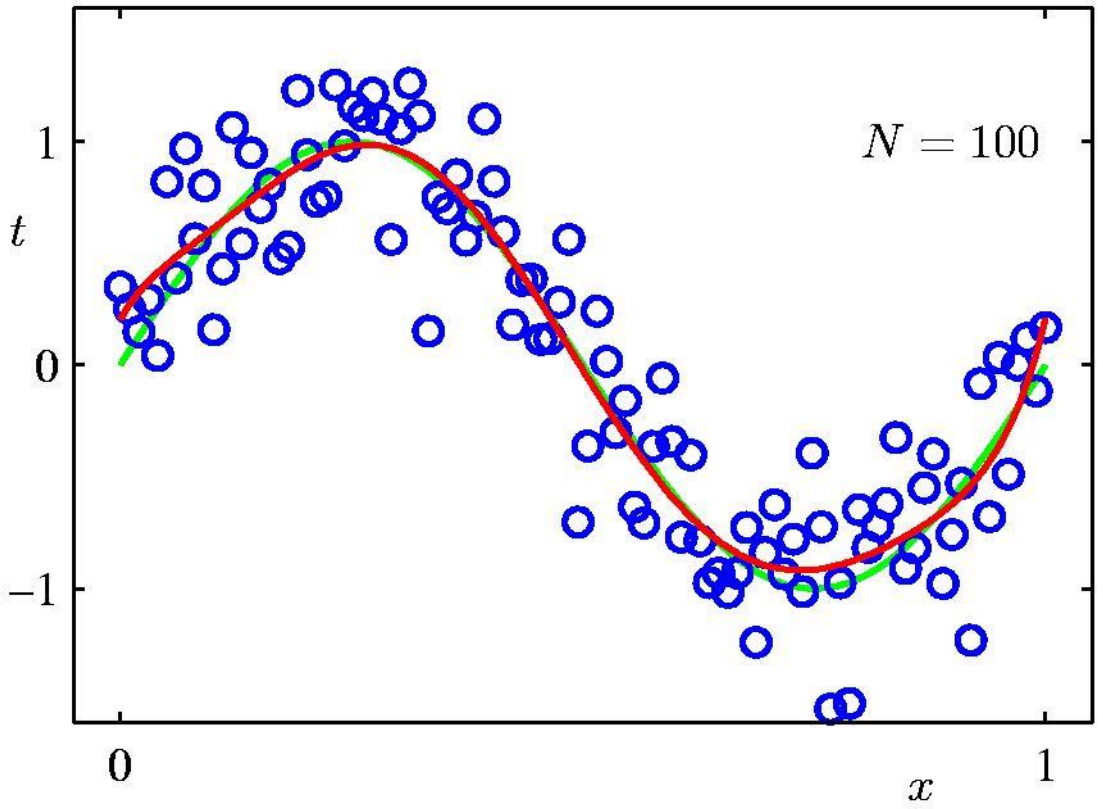
- Pattern Recognition and Machine Learning, *Christopher Bishop*



- Lecture notes of *Andrew Ng*, Stanford CS229

Regression

- Similar to curve fitting



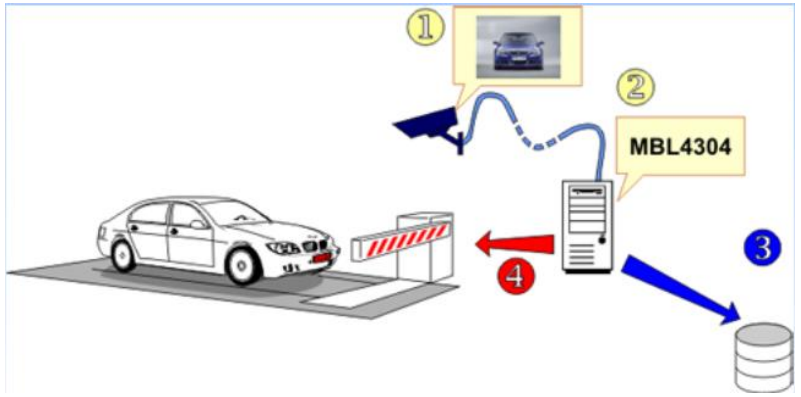
Classification

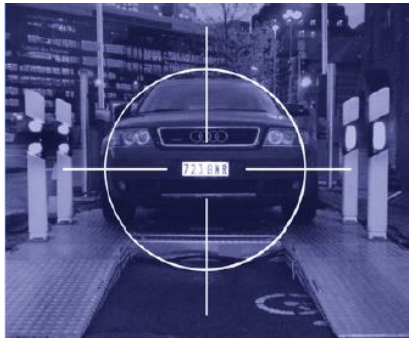
- 2 classes
- Labeled data

Courtesy of Anil Jain

License Plate Reading System

- Detect and read the license plates





- Modules:
  - acquisition,
  - enhancement,
  - segmentation,
  - character recognition
- Accuracy, robustness & real-time

Courtesy of Anil Jain

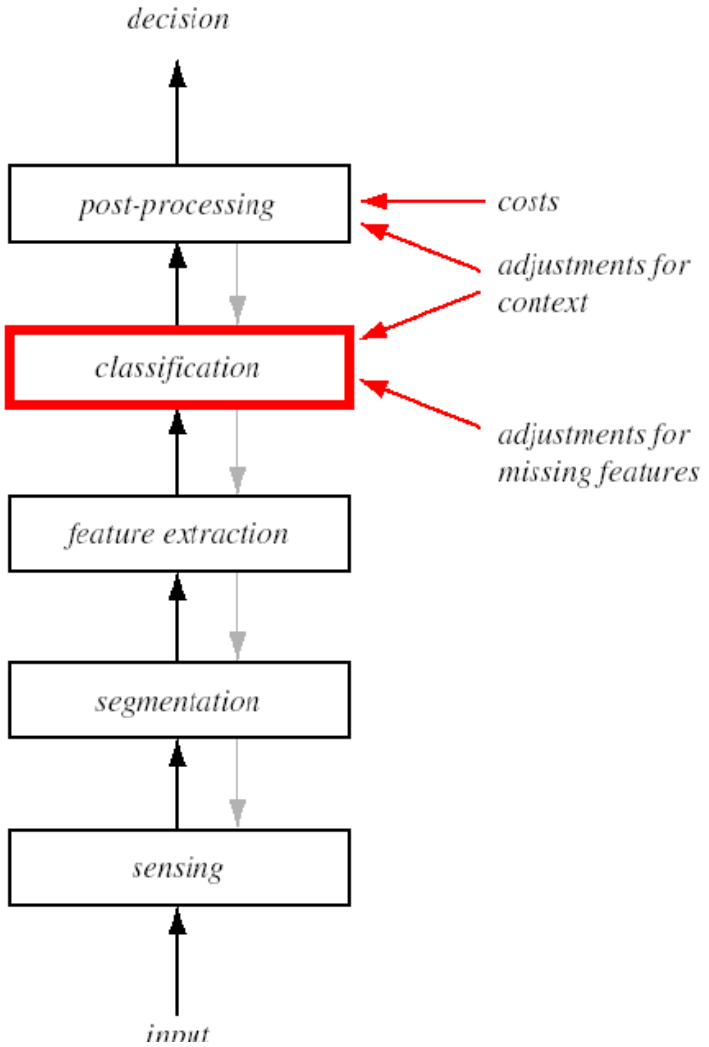
Processing steps

- Plate localization: Isolate the plate in image
- Preprocessing: Plate orientation and sizing
- Normalization: Adjust image brightness & contrast
- Segmentation: Find individual characters
- Character recognition: OCR
- Post-processing: Rules for character placement

Challenges

- Poor image resolution: plate too far; low-res. image
- Motion blur
- Low contrast: overexposure, reflection/shadows
- Viewpoint variation and occlusion
- Different fonts, background

Classification



Courtesy of Anil Jain

Representation

Desirable properties

- Invariance
- Account for intra-class variations
- Ability to discriminate classes of interest; low inter-class similarity
- Robustness to noise, occlusion,..
- Provide simple decision making strategies
- Low measurement cost; real-time

Invariant representation

- Invariant to
  - Translation
  - Rotation
  - Scale
  - Skew
  - Deformation
  - Color

Not all invariant properties are needed for a given application

Fish sorting

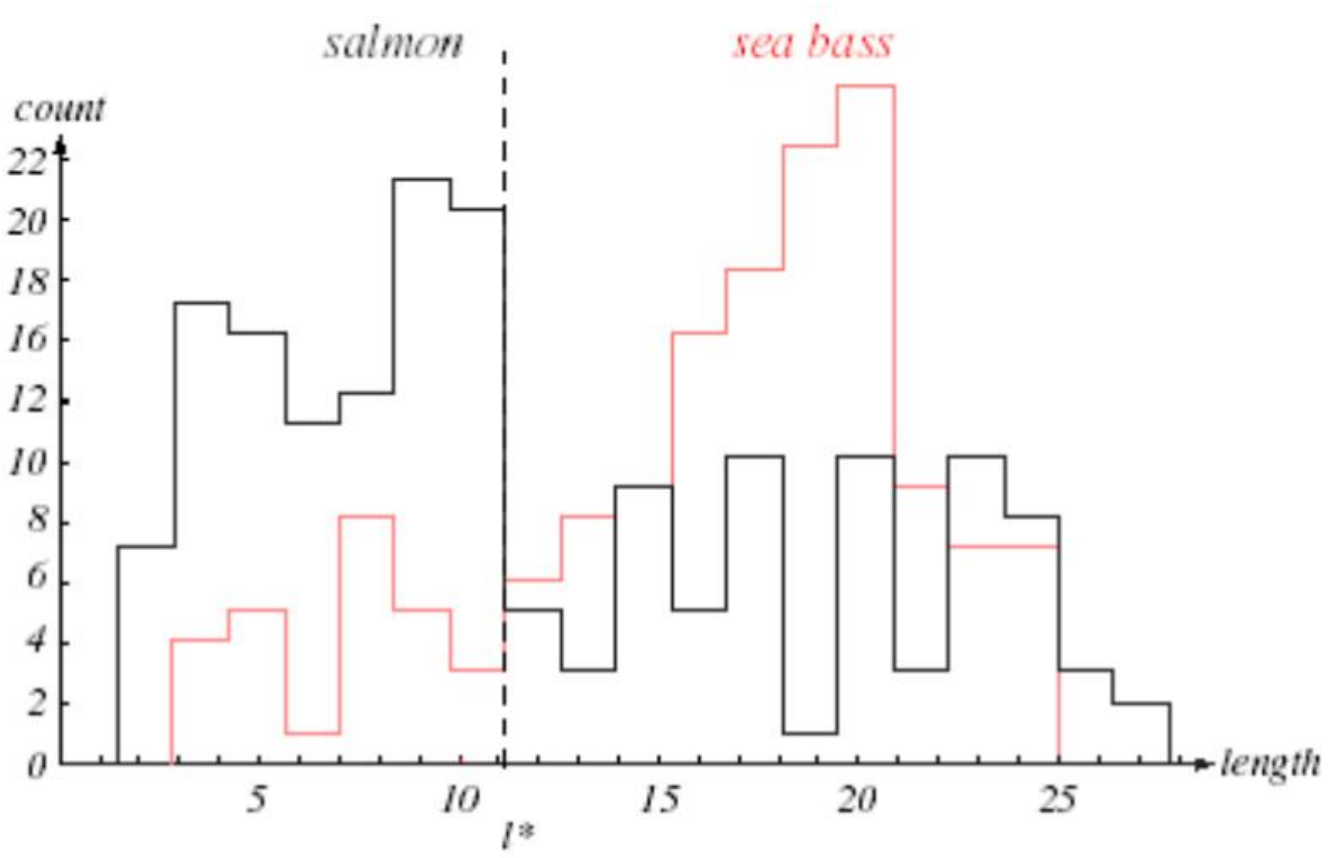


FIGURE 1.2. Histograms for the length feature for the two categories. No single threshold value of the length will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value marked *l** will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Fish sorting

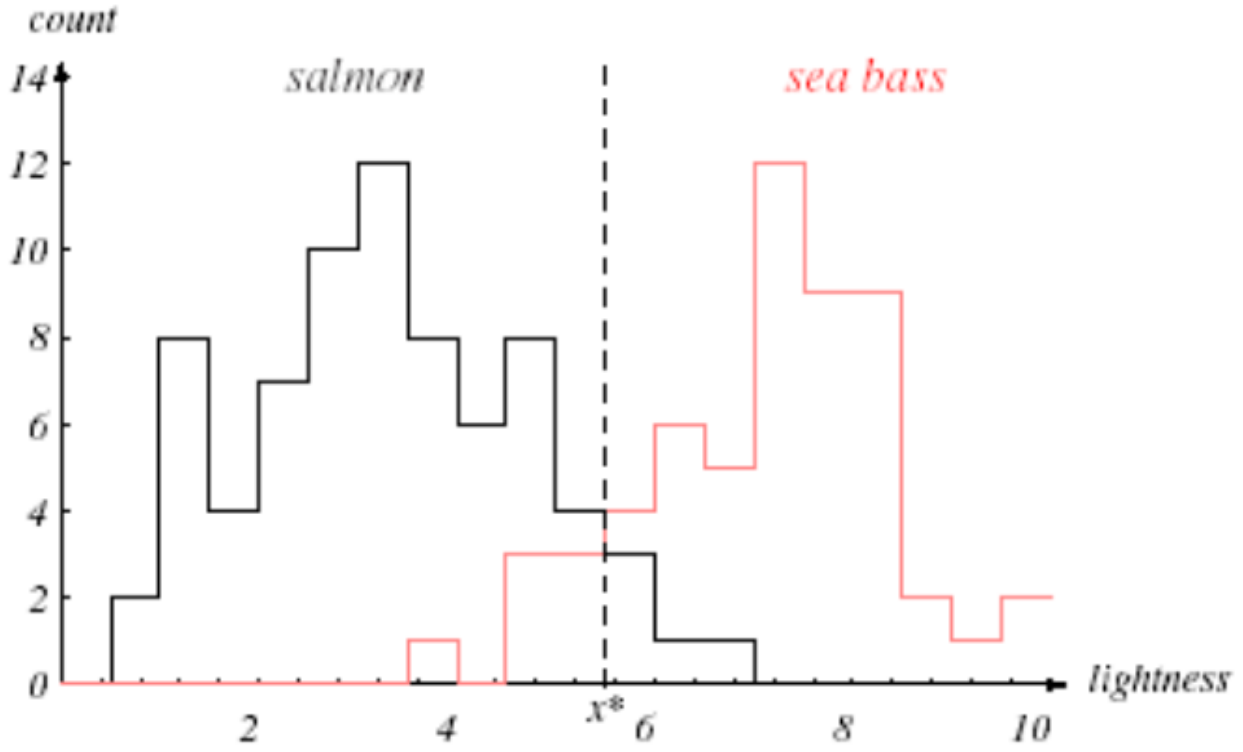


FIGURE 1.3. Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Overlap of these histograms is small compared to length feature

Two-dimensional Feature Space

Linear decision boundary; linear classifier

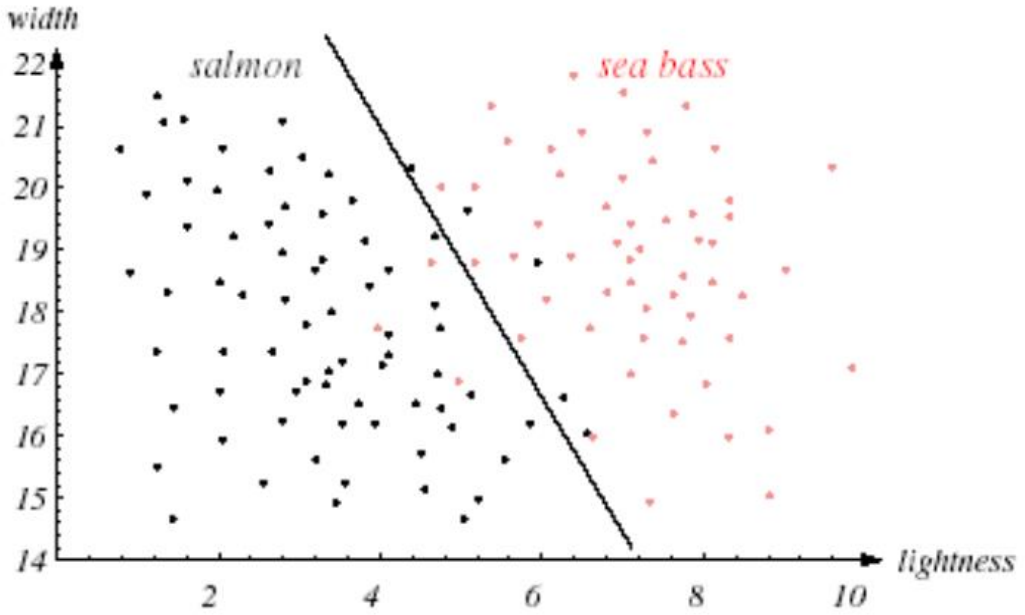


FIGURE 1.4. The two features of lightness and width for sea bass and salmon. The dark line could serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Joint distribution of two features leads to better separation

Complex decision boundary

Polynomial classifier

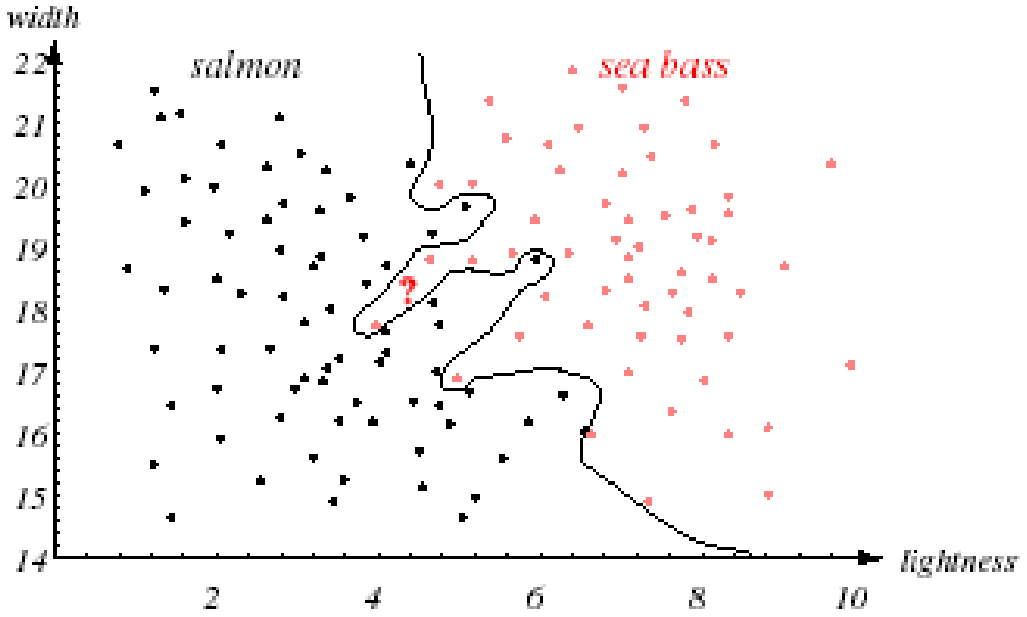


FIGURE 1.5. Overly complex models for the fish will lead to decision boundaries that are complicated. While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns. The novel test point marked ? is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be classified as a sea bass. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

What is the generalization ability of the classifier?

Good Generalization & Good Accuracy

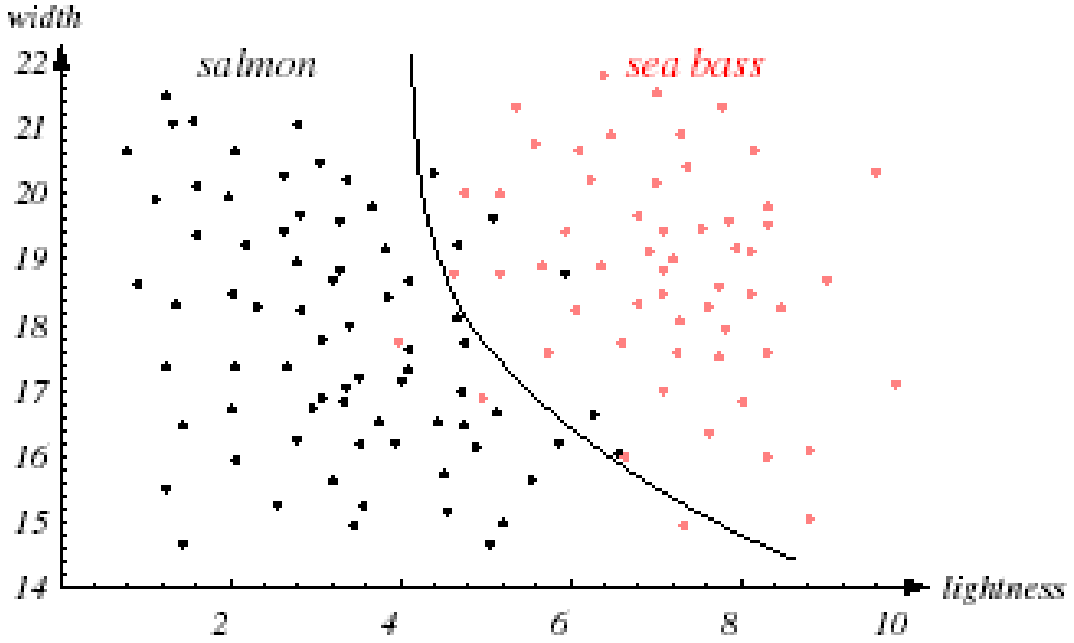


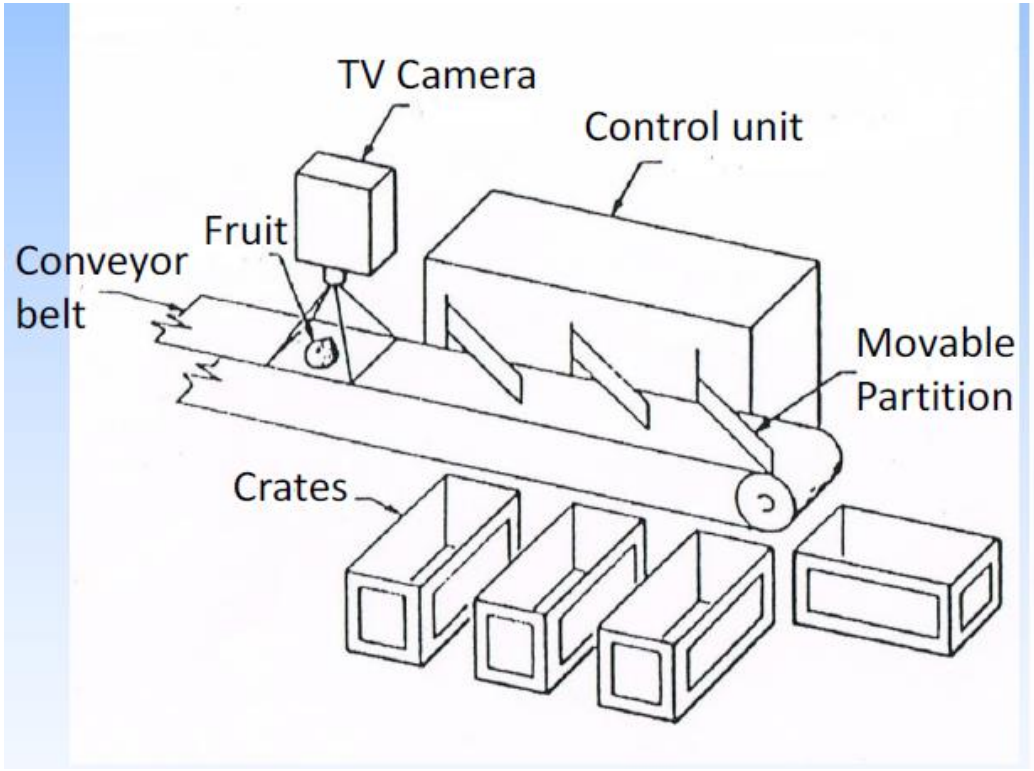
FIGURE 1.6. The decision boundary shown might represent the optimal tradeoff between performance on the training set and simplicity of classifier, thereby giving the highest accuracy on new patterns. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

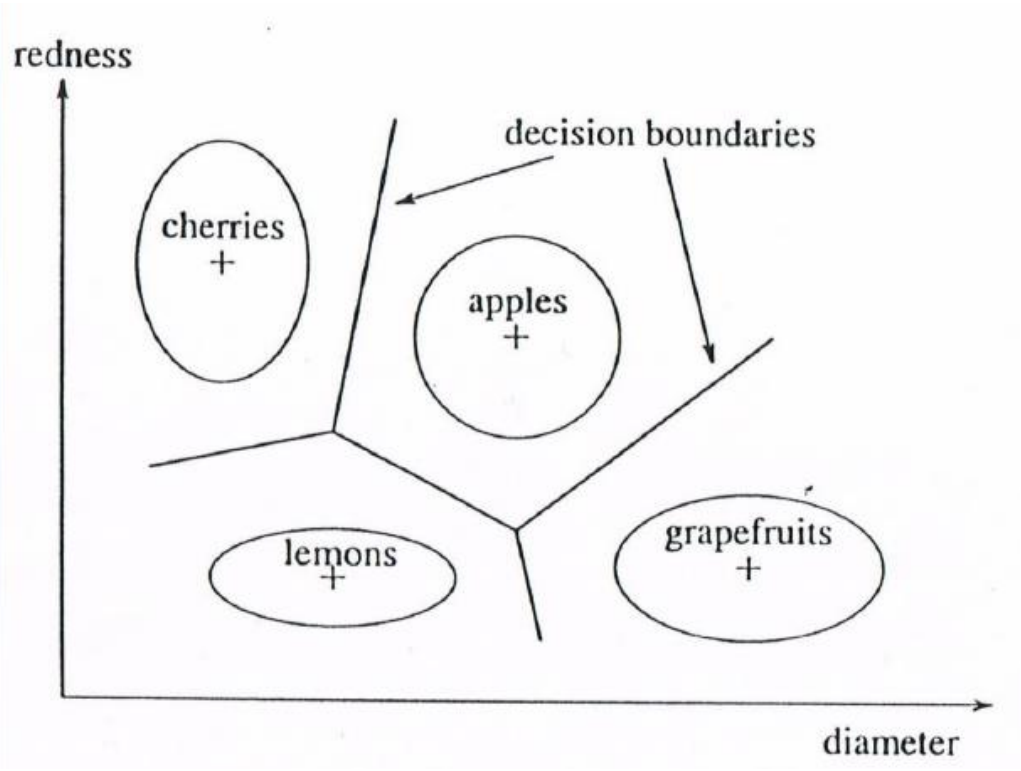
Simple decision boundaries are preferred

Occam’s Razor (William of Ockham (c. 1287–1347)

"If you have two equally likely solutions to a problem, choose the simplest"

Fruit sorter





Reject option

- What if the system encounters a previously unseen class?

ض

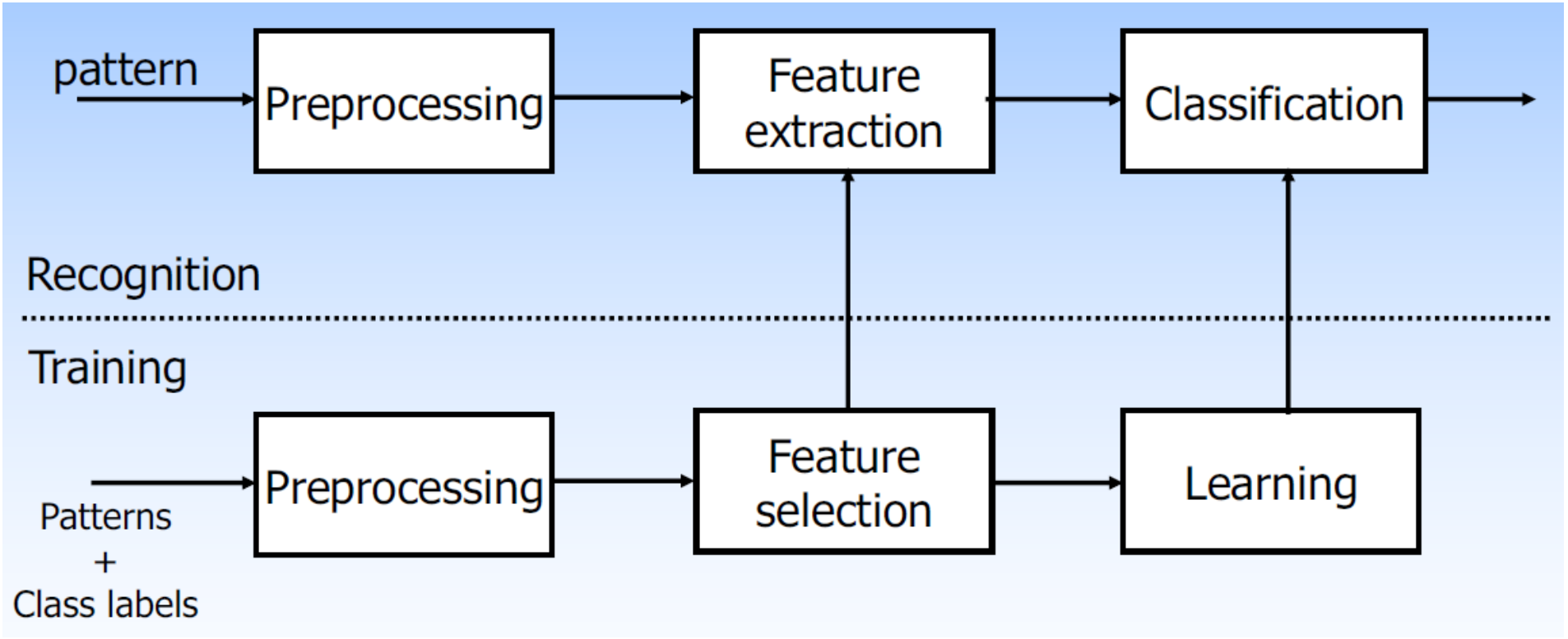
Feature Selection & Extraction

- Feature selection: which subset to use? Some features may be redundant

- Feature extraction: which combination of given features to use?

- Curse of dimensionality:Error rate may in fact increase with too many features in the case of small number of training samples

Statistical Pattern Recognition

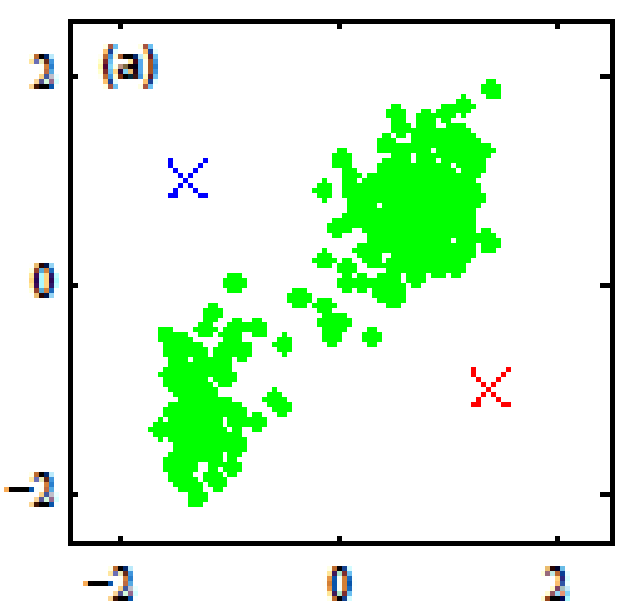


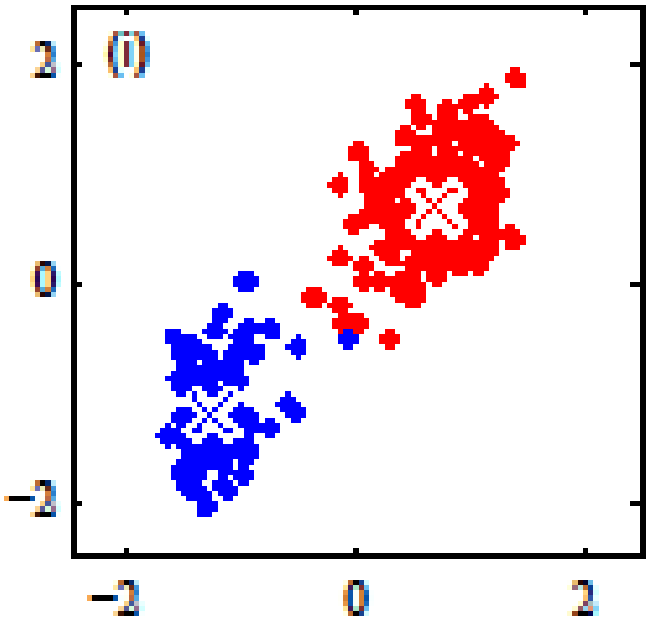
Clustering

- Unlabeled data
- How many clusters??

Courtesy of Anil Jain

K-means Clustering





Reinforcement Learning

- RL demo 1: https://youtu.be/fiQsmdwEGT8
- RL demo 2: https://youtu.be/bPkWMlCq2tc

Grade distribution

- Coursework (HWs/assignments) 20 points
- Final 80 points