Statistical Methods in Artificial Intelligence
CSE471 - Monsoon 2016: Lecture 01

Avinash Sharma

CVIT, IIIT Hyderabad

Course Content

- Introduction
- Linear Classification
- Neural Networks
- Probability Densities
- Bayesian Classifiers
- Dimensionality Reduction
- Support Vector Machines
- Kernel Methods
- Clustering Techniques
- Decision Tree/Graphical Models

Reference Material

- Books
  - ❖ **“Pattern Classification” by Duda, Hart & Stork**
  - ❖ “The Elements of Statistical Learning” by Hastie, Tibshirani and Friedman
  - ❖ “Machine Learning : A probabilistic Perspective” by Kevin P. Murphy
- Pre-requisite
  - ❖ Basics of Linear Algebra, Calculus, Probability Theory and Statistics. Programming in Matlab and C/C++.
- Course Website http://moodle.iiit.ac.in
- Online Courses/Tutorials and Research Papers

Expected Outcome

- This course would enable students to understand pattern recognition techniques in detail.
- We will ensure that both ***theoretical*** as well as ***practical*** aspects are learnt simultaneously.
- The project deliverables are expected to working code/prototype where sound theoretical frameworks are employed or extended for practical applications.

Grading Scheme

- Relative Grading.

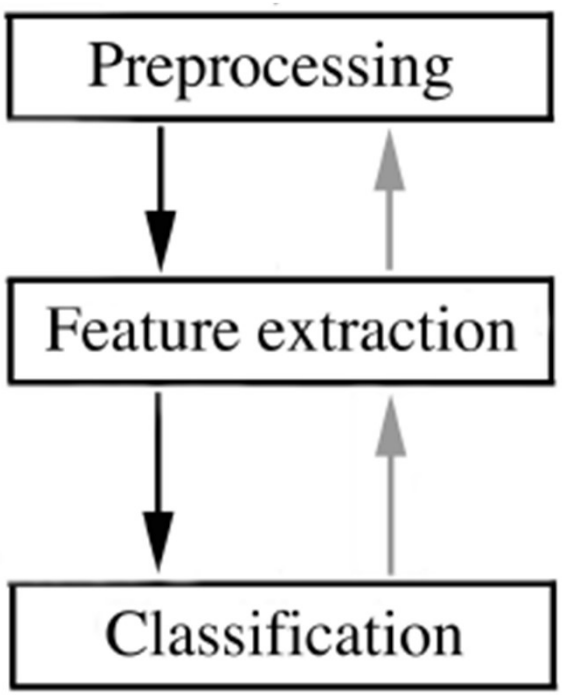
- Detailed Grading Breakup (~ 60% Exams/Quiz + 40% Practical Learning):
  - 24% Course Project
  - 16% Assignments (4×4)

  - 26% Mid-semester Exams (13×2)
  - 30% Final Exam
  - 04% Quiz/Others

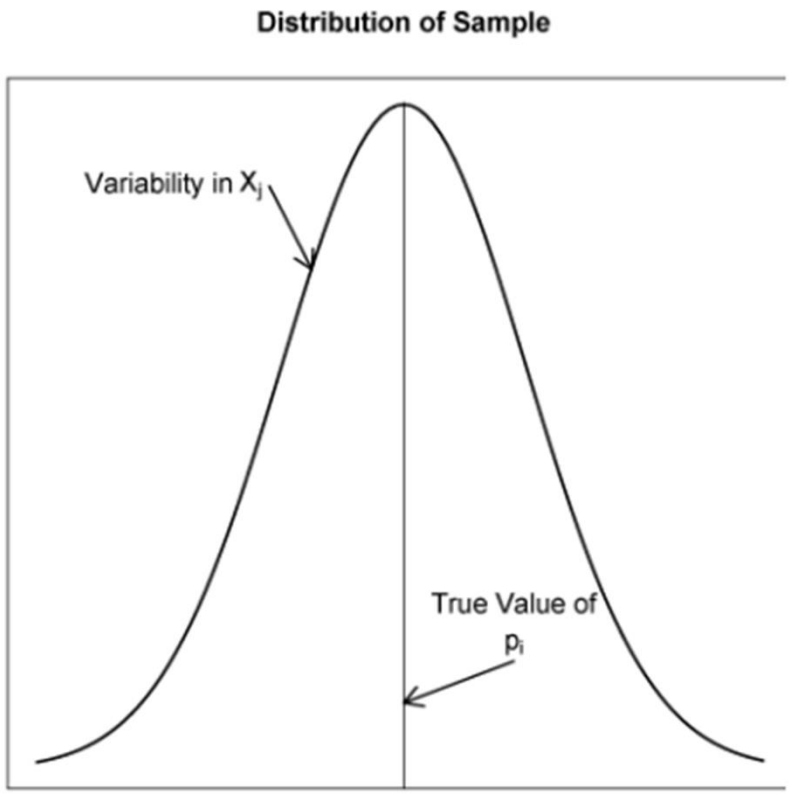
- Zero marks in assignment if a copying is detected
- Cooperation with TA's is expected
- Zero tolerance from institute's attendance policy

Introduction

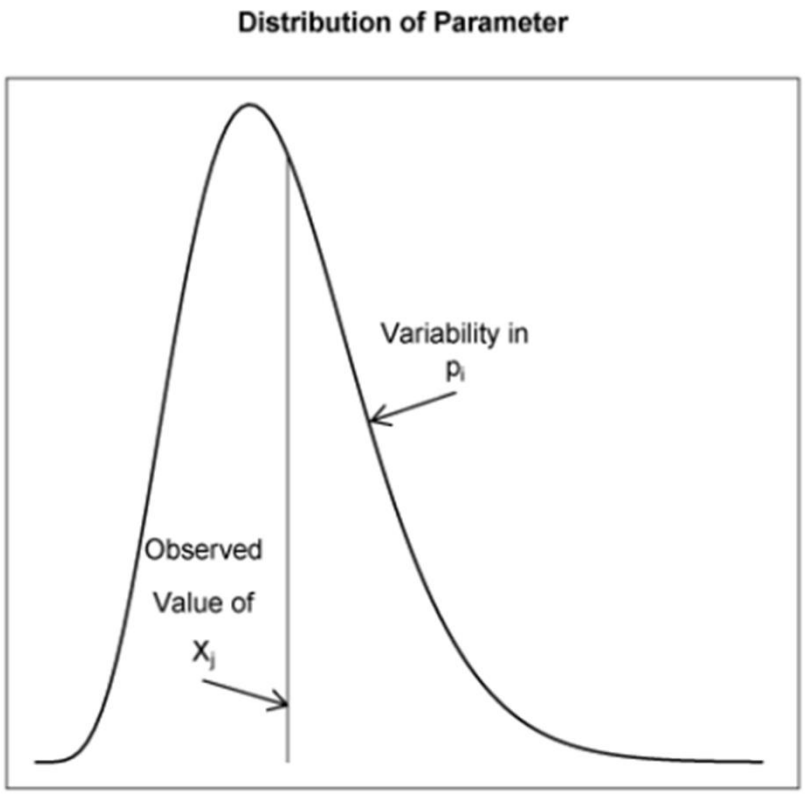
- Statistical Methods
  - Frequentist Approach
  - Bayesian Approach
- Artificial Intelligence
  - Pattern Recognition (PR)
- Our focus is “SM in AI”
- Key Problems in PR
  - **Classification** (Assignment to category labels)
  - Clustering (Seeking natural division of data)



Introduction



Frequentist Approach



Bayesian Approach

Classification

- Preprocessing: It involves cleaning the input data for noise and clutter.

- Feature Extraction:
  - Each property/attribute of a data instance/observation/sample is termed a feature, also known in statistics as an *explanatory variable*.

  - A *feature vector* is typically a multi-dimensional vector that stores ***individual*** *and* ***measurable*** properties of each of data instance.

  - Features can be Binary, Categorical or Real-valued.

Classification

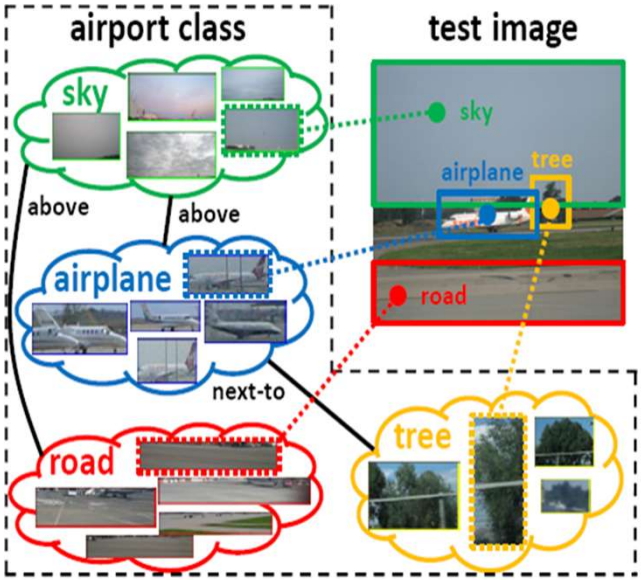
- Identification of a specific category label from a set of categories a new observation in the test set of data belongs to.

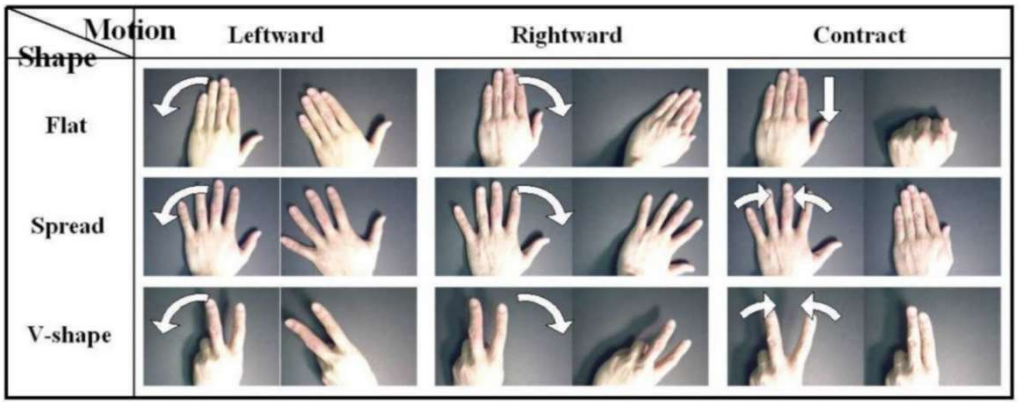
- This is achieved using the training set of data containing observations whose category membership is already known.

- Both training and test set of data samples are assumed to be *drawn* form the ***same distribution***.

Applications

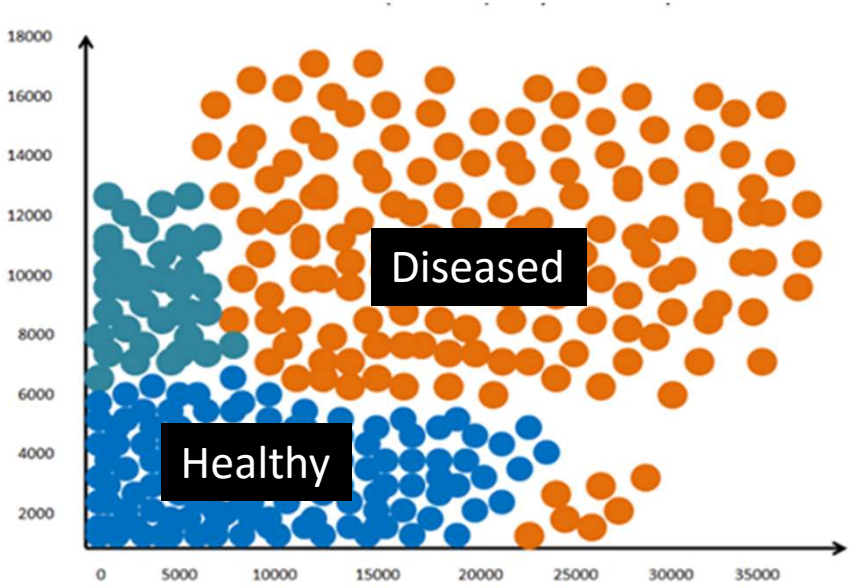
- **Vision/Robotics**
- Healthcare
- Life Sciences
- Text Analytics
- Speech Processing
- Business
- Surveillance
- …

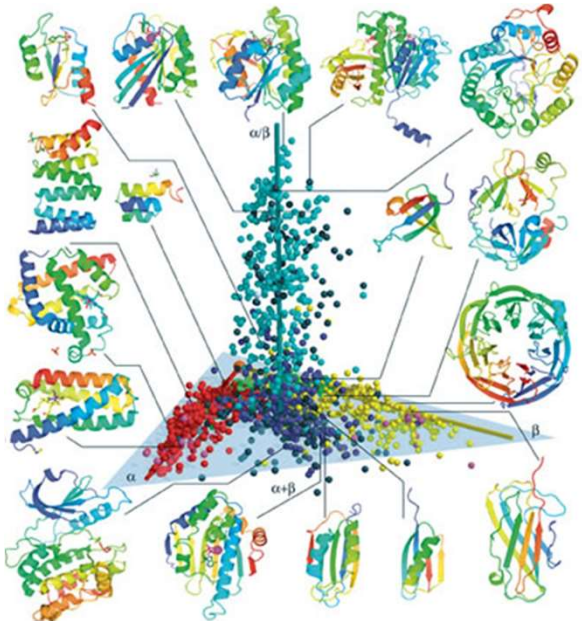




Applications

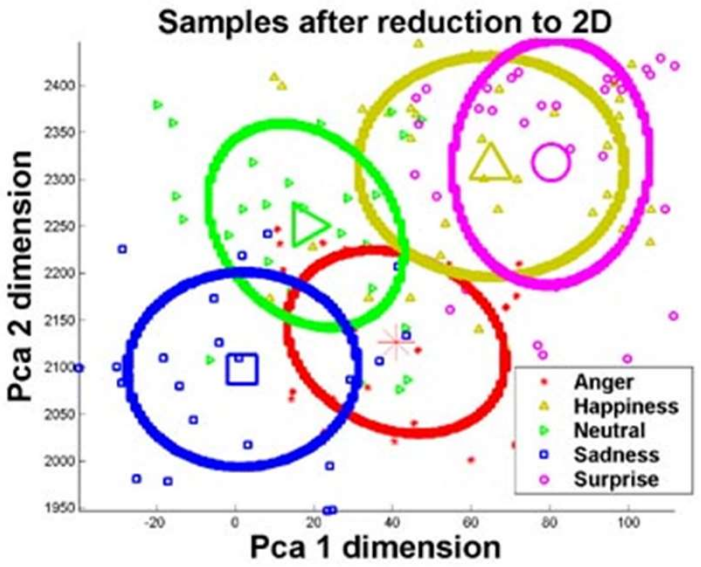
- Vision/Robotics
- **Healthcare**
- **Life Sciences**
- Text Analytics
- Speech Processing
- Business
- Surveillance
- …





Applications

- Vision/Robotics
- Healthcare
- Life Sciences
- **Text Analytics**
- **Speech Processing**
- Business
- Surveillance
- …



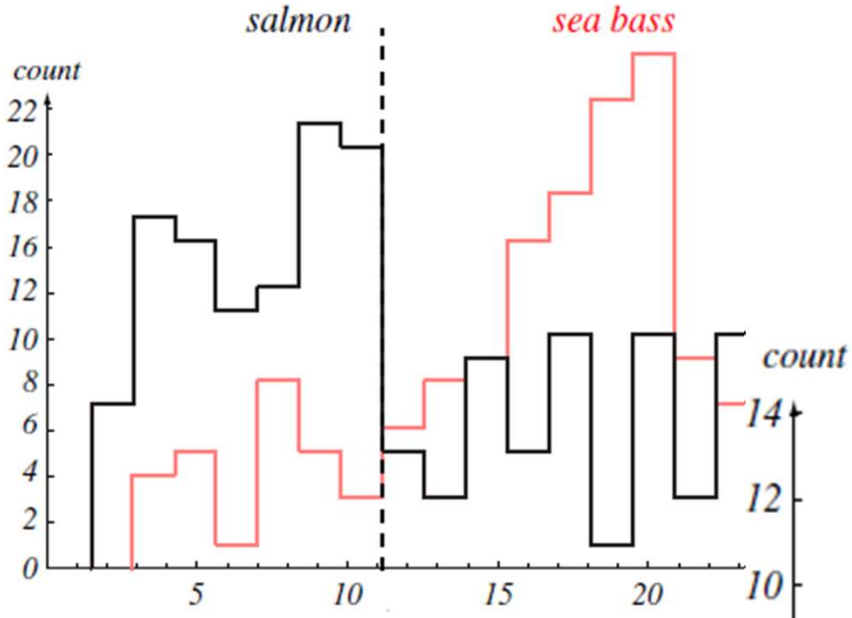
Toy Examples Walkthrough

- Image Based Fish Classification (Salmon v/s Sea Bass)

Toy Examples Walkthrough

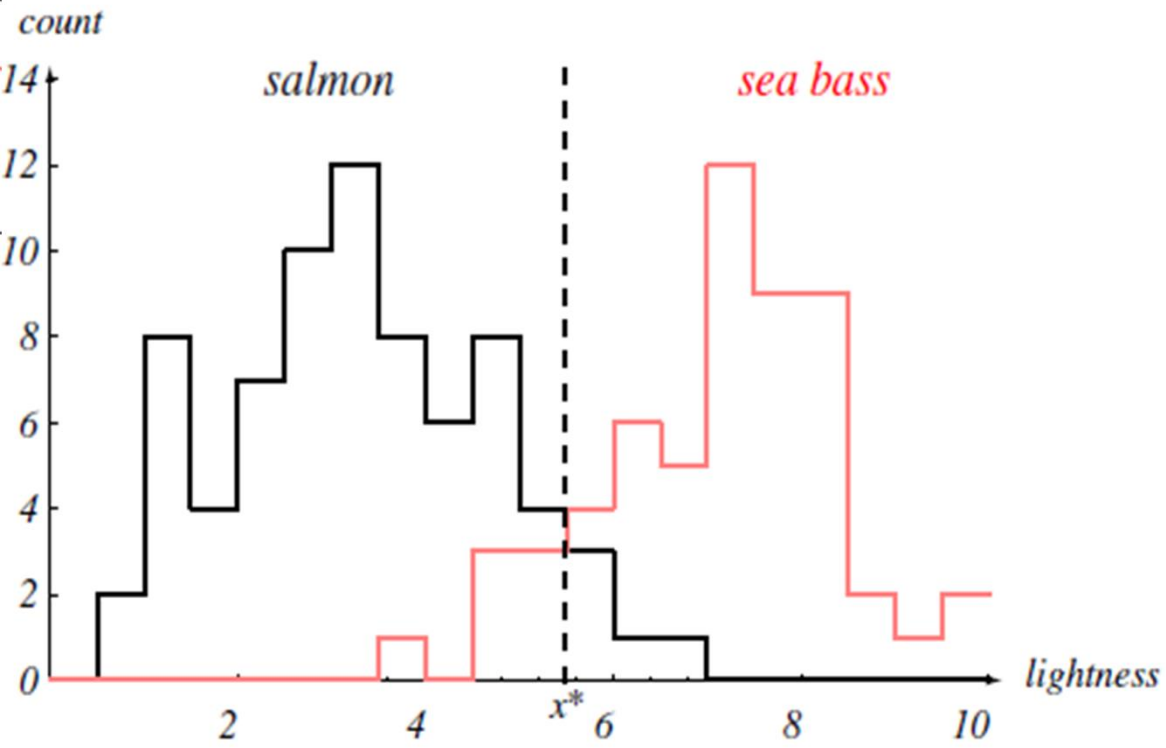
- What could be the best **features** representing these two fish?
  - Length
  - Width
  - Appearance
    - Lightness (reflective or shiny skin)
    - Texture
  - Fin Structure
  - …

Toy Examples Walkthrough



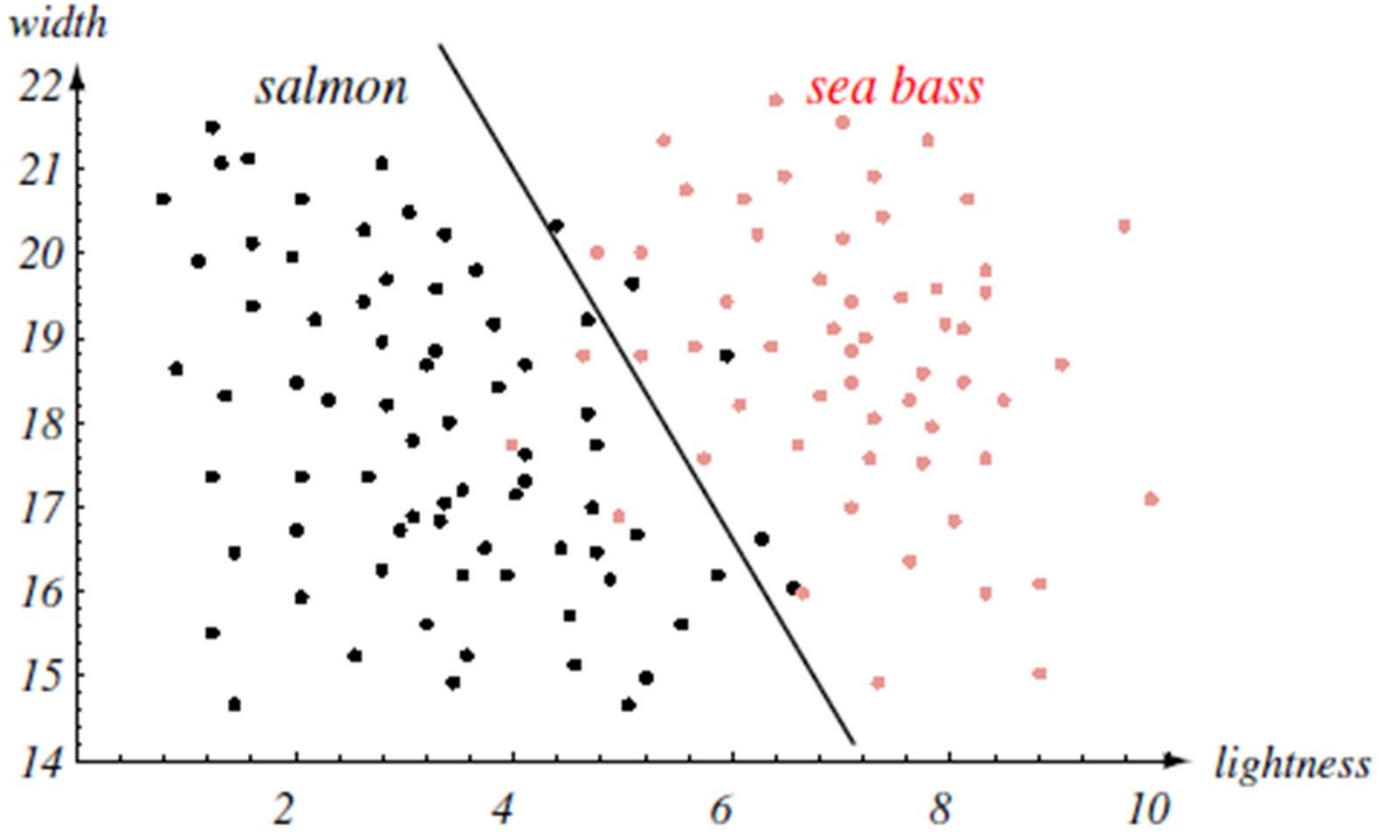
Length Histogram Feature

Classification Cost
&
Decision Theory

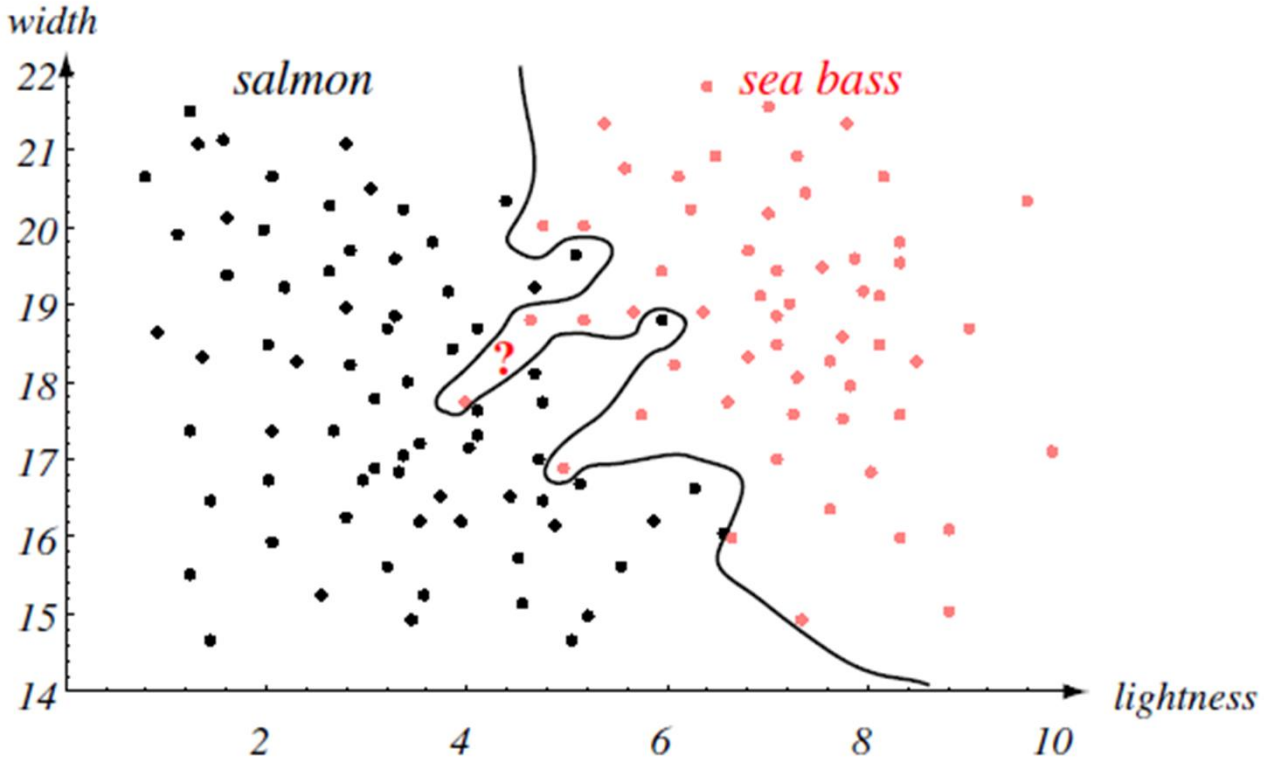


Lightness Histogram Feature

Toy Examples Walkthrough



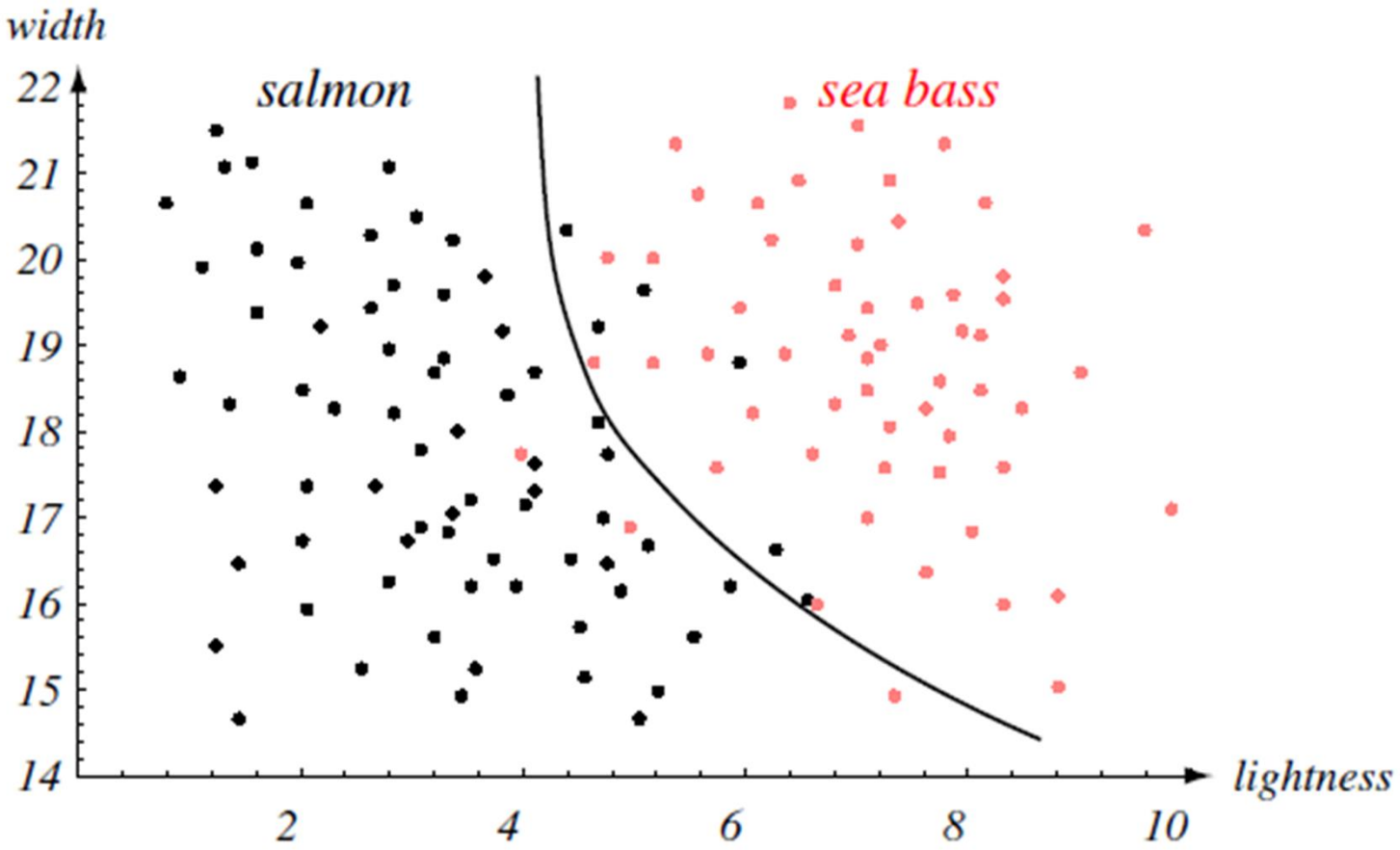
Toy Examples Walkthrough



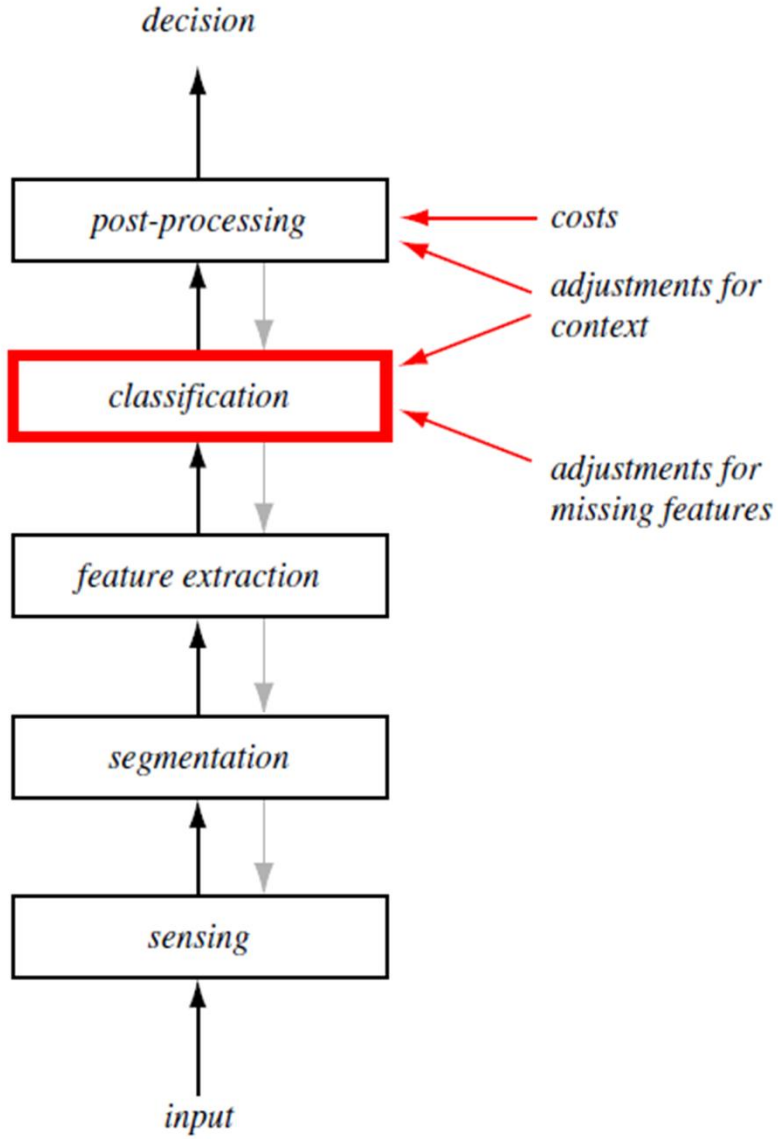
- **Generalization v/s Overfitting**

- **Penalizing complex models** : “The principle of Occam Razor states that among competing hypotheses that predict equally well, the one with the fewest assumptions should be selected. “

Toy Examples Walkthrough



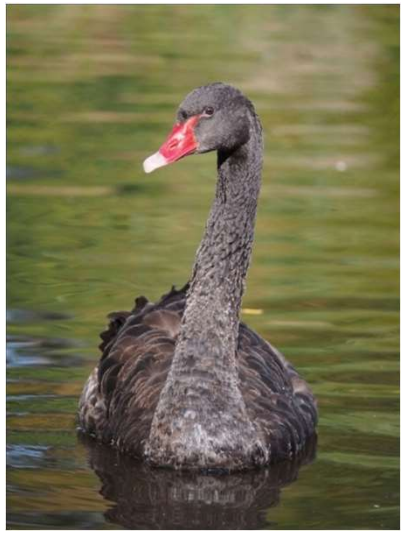
PR System Flow



Challenges

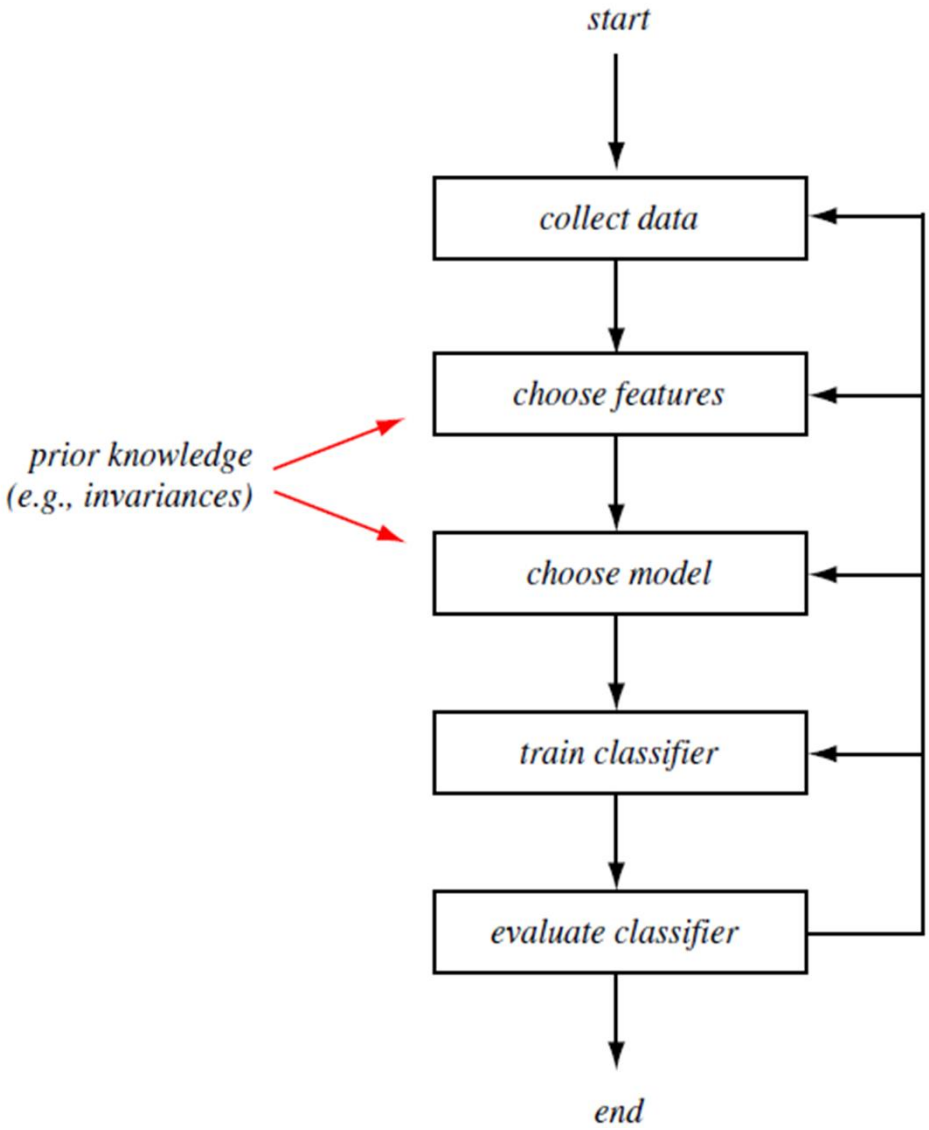
- Segmentation
  - Recovering the Signal from noise/clutter
- Feature Extraction
  - Invariance to Translation, Rotation and Scaling
  - Occlusion
  - Rate
  - Deformation
  - Selection of best features

Challenges



- Classification
  - Noise “Any property of the sensed pattern which is not due to true underlying model but instead to randomness in the world or the sensor”.
- Post Processing
  - Error Rate
  - Risk
  - Context
  - Multiple Classifiers

PR System Design Cycle



Types of Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning