Introduction to Pattern Classification/Recognition

Chapter 1 (Duda et al.)

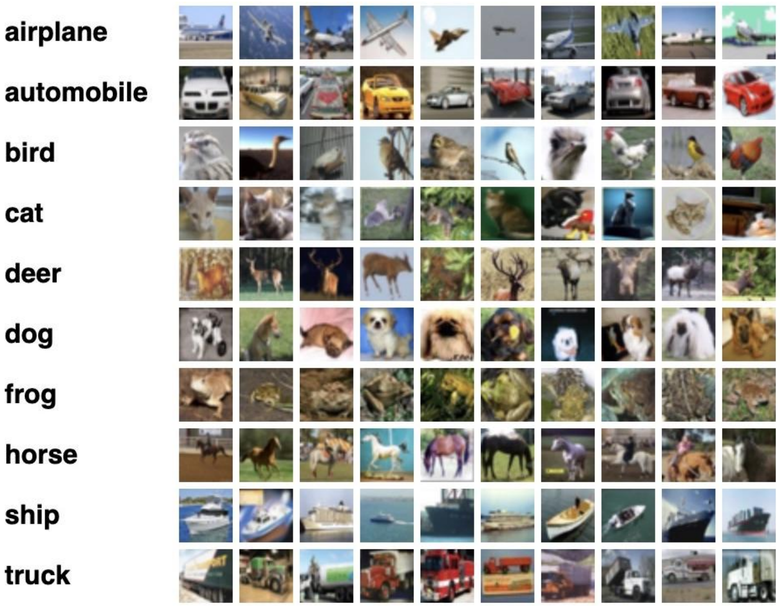
CS479/679 Pattern Recognition
Dr. George Bebis

What is a Pattern?

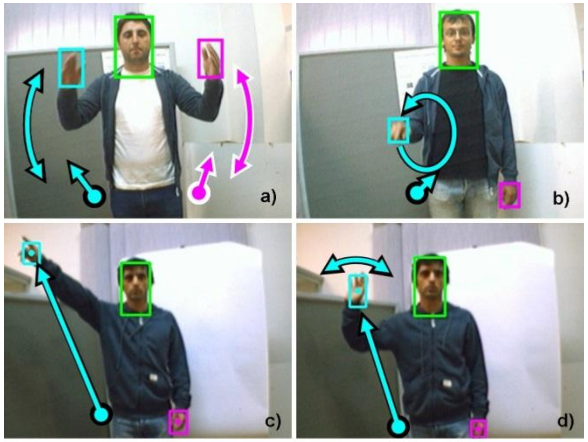
- An object or event that has some structure or regularity.
- Typically, represented by a vector **x**, of D "features" x_i.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_D \end{bmatrix}$$

Objects



Events (e.g., hand gestures)



What is a Class ?

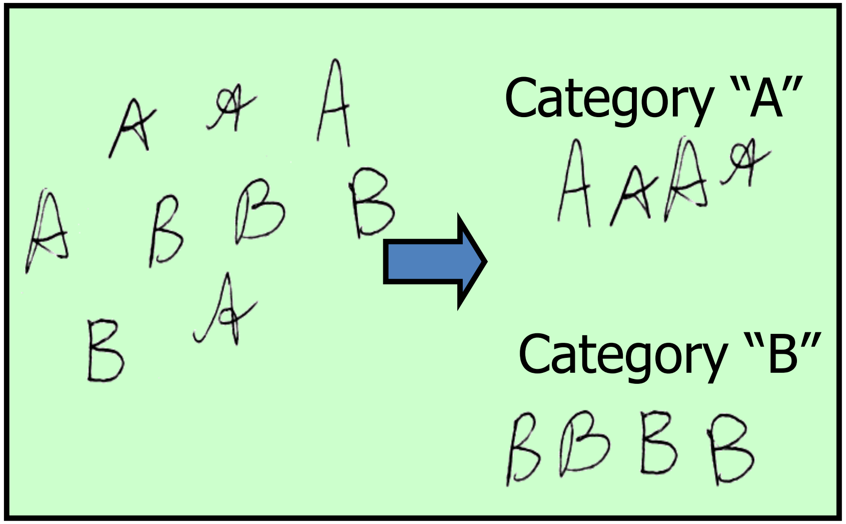
- A collection of "similar" objects (based on appearance or semantics).

Dog class

Cat class

Pattern Recognition / Classification

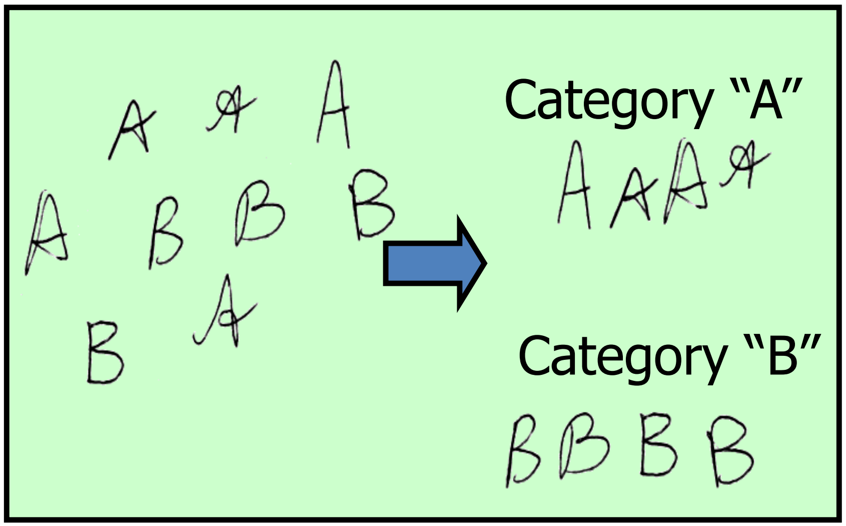
- Assign a pattern to one of several known categories or classes.



Classification vs Clustering

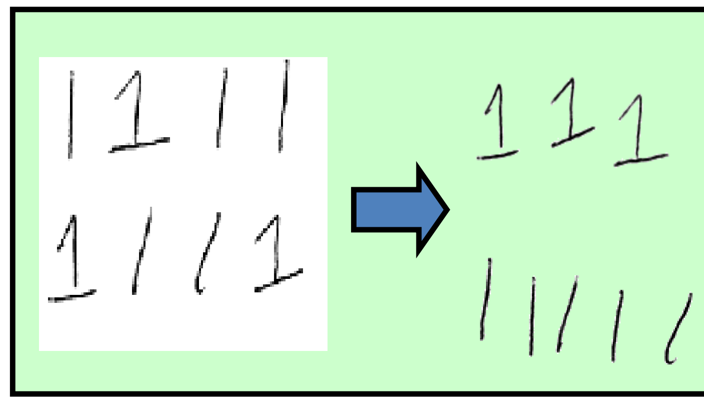
Classification: known categories

(Supervised Classification)



Clustering: unknown categories

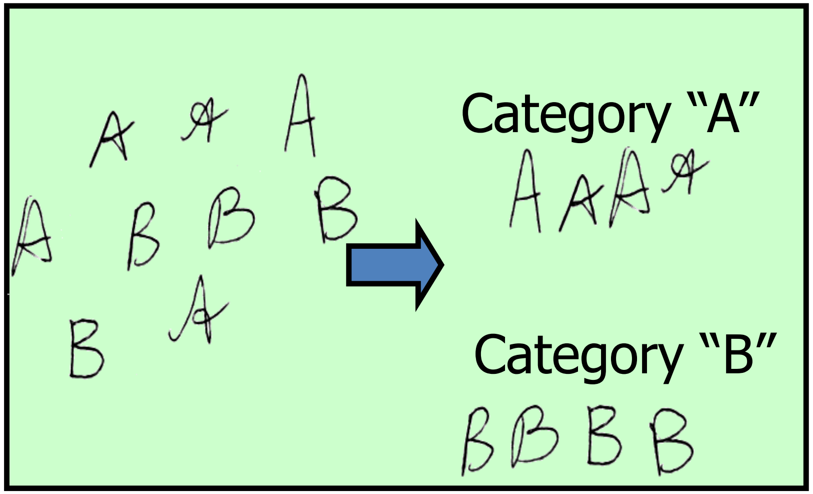
(Unsupervised Classification)



Classification vs Regression

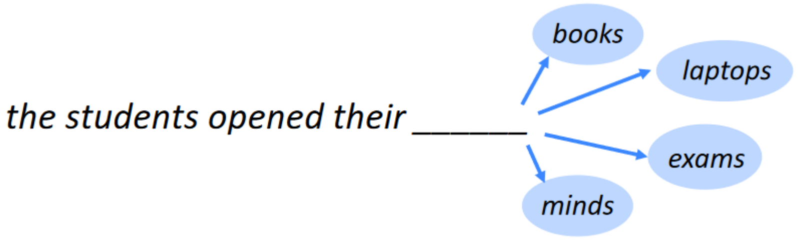
Classification: known categories

(Supervised Classification)



Regression: predict one or more output variables (i.e., dependent variables) from one or more input input variables (i.e., independent variables).

e.g., next-word prediction

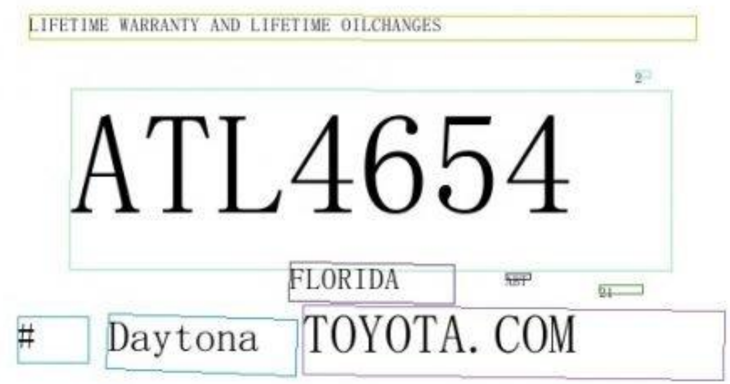


Applications

Character Recognition

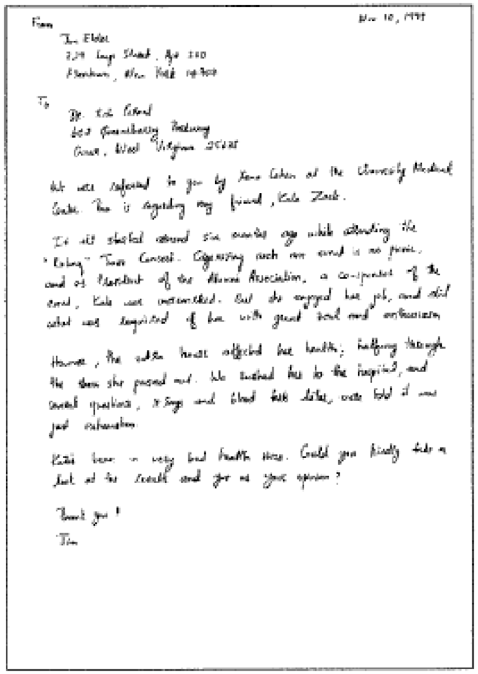
(machine-printed)

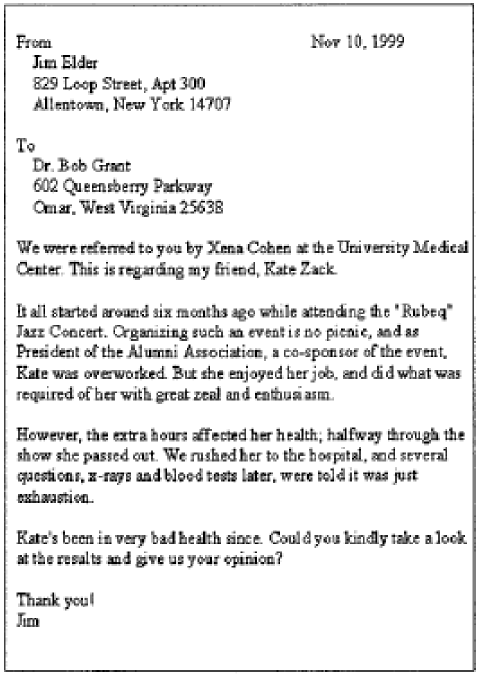
License Plate Recognition



Character Recognition
(handwritten)

Document Analysis





From Nov 10, 1999
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

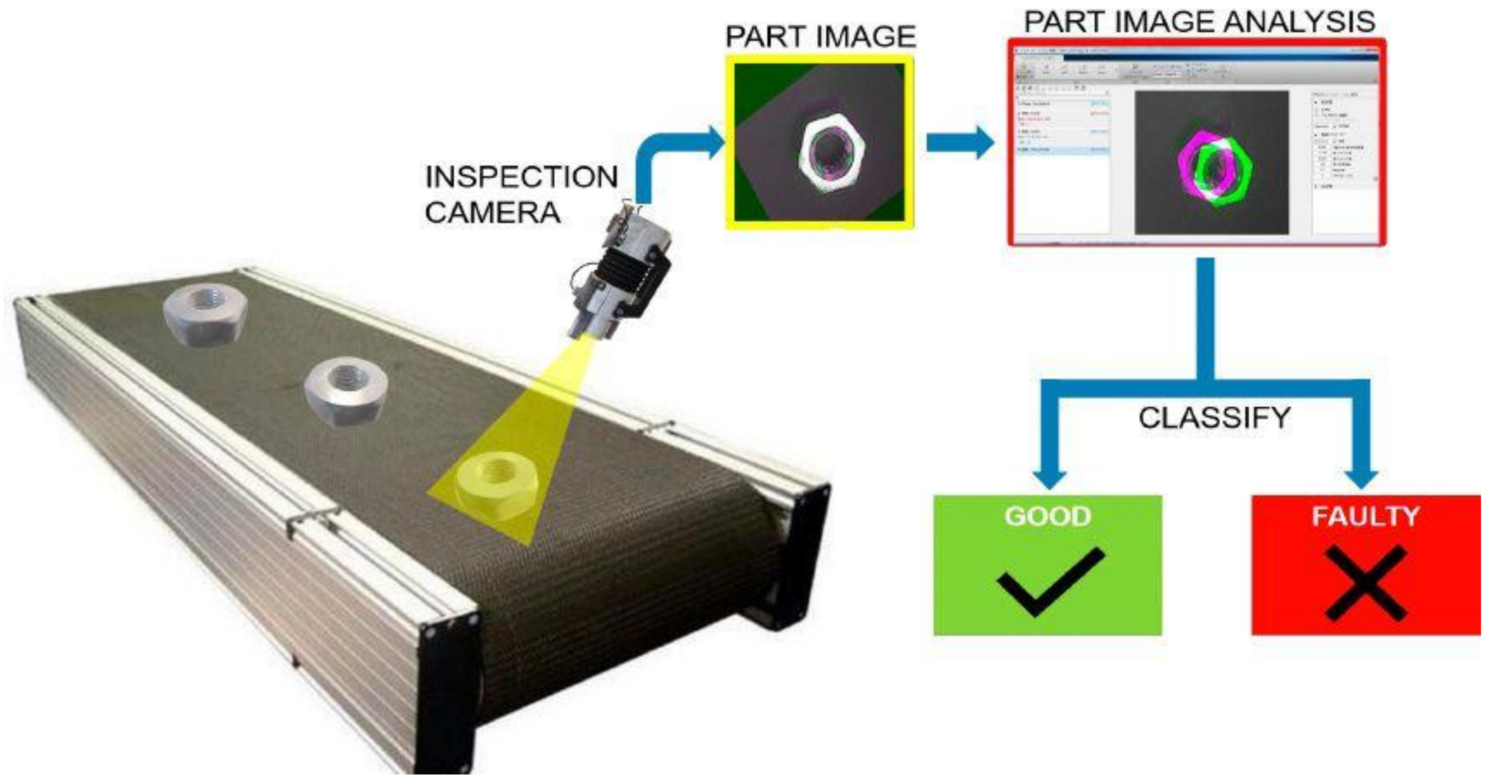
It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

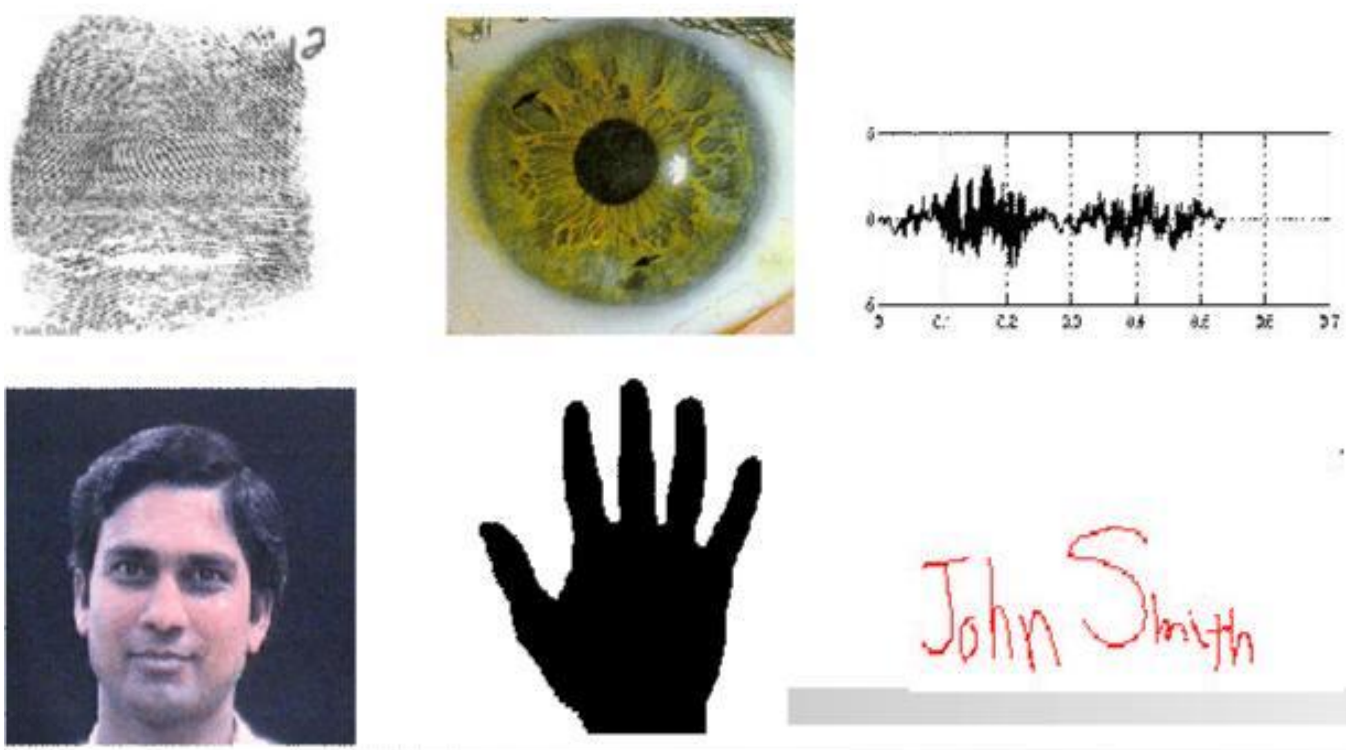
Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

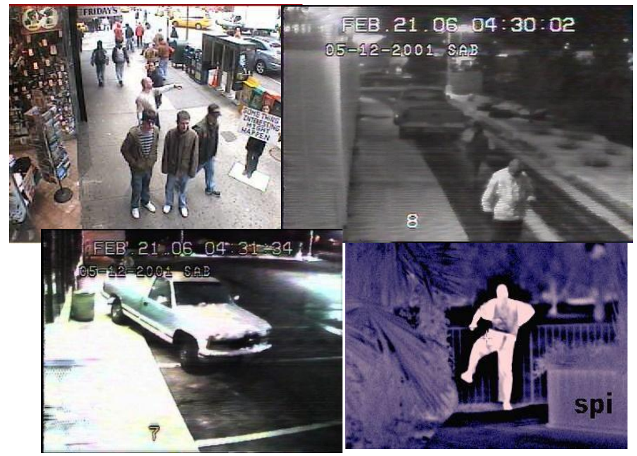
Visual Inspection

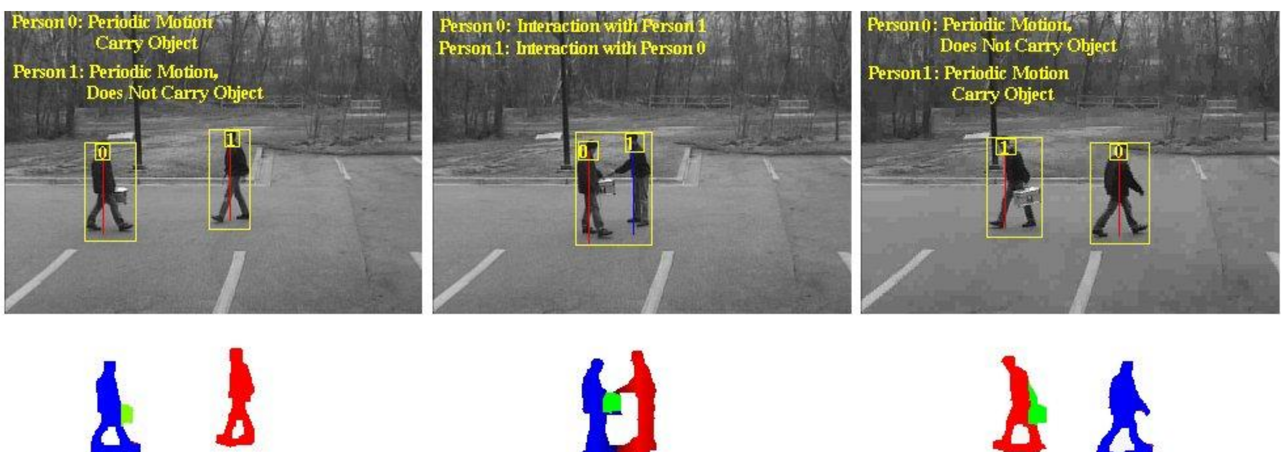


Biometric Recognition



Visual Surveillance & Human Activity Recognition

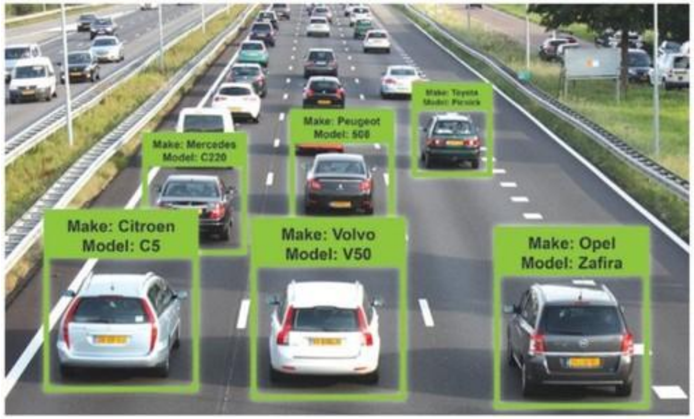




Traffic Monitoring

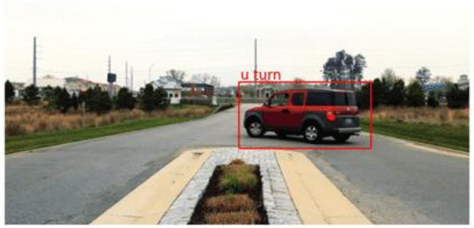
Anomaly Detection

Vehicle Classification



traffic congestion

car in wrong way

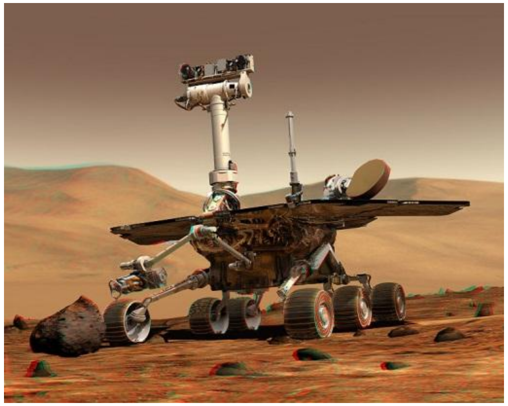


risky u-turn

traffic collision

Autonomous Systems

Object Recognition
Obstacle Detection and Avoidance



Wildfire Monitoring

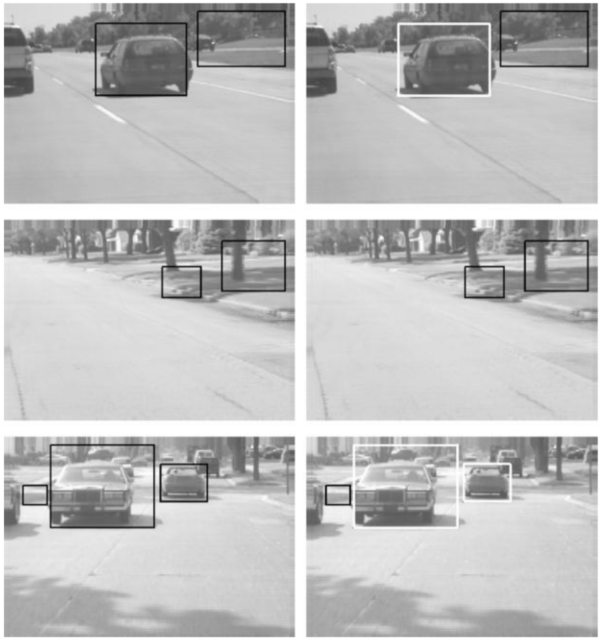
Fire Detection and Tracking

Other Applications

- Recommendation systems
  - e.g., Amazon, Netflix
- Loan/Credit Card Applications
- Spam filters
- Malicious website detection

Related Work @ CVL
(Computer Vision Lab)

Pre-crash Vehicle Detection



Z. Sun, G. Bebis, and R. Miller, "Monocular Pre-crash Vehicle Detection: Features and Classifiers", **IEEE Transactions on Image Processing**, vol. 15, no. 7, pp. 2019-2034, July 2006.

Related Work @ CVL

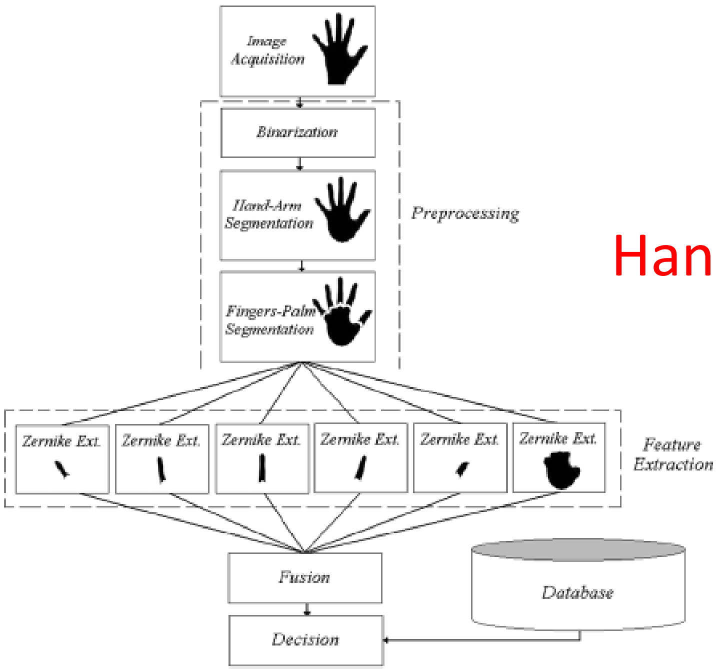


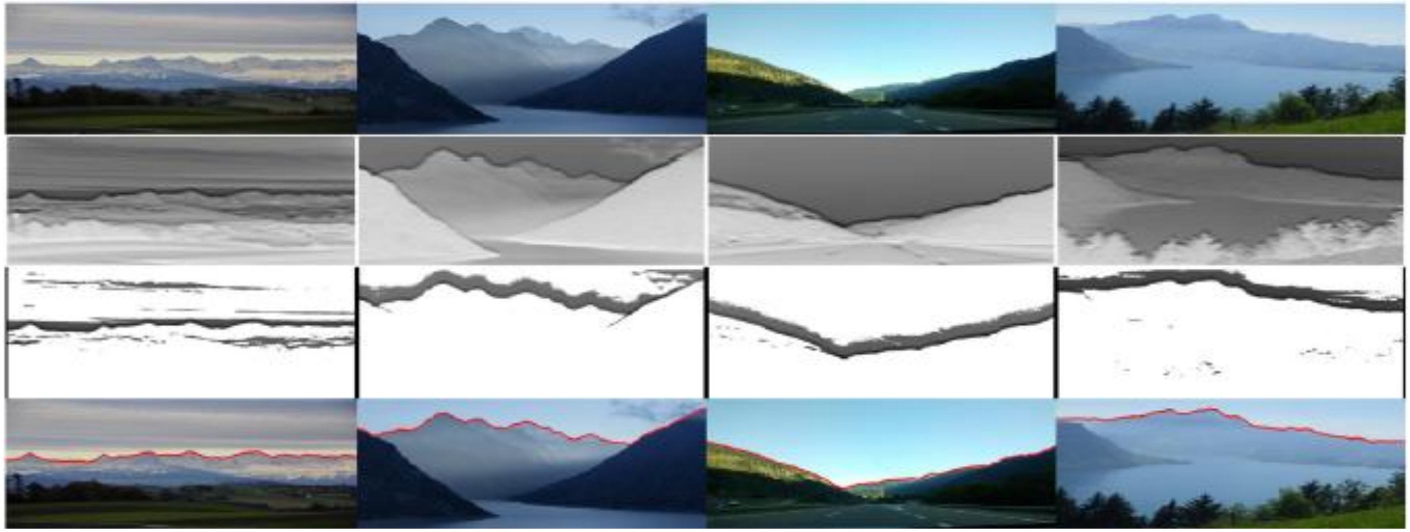
Fig. 1. Main stages of the proposed system.

Hand-based Verification

G. Amayeh, G. Bebis, A. Erol, and M. Nicolescu, "Hand-Based Verification and Identification Using Palm-Finger Segmentation and Fusion", **Computer Vision and Image Understanding (CVIU)** vol 113, pp. 477-501, 2009.

Related Work @ CVL

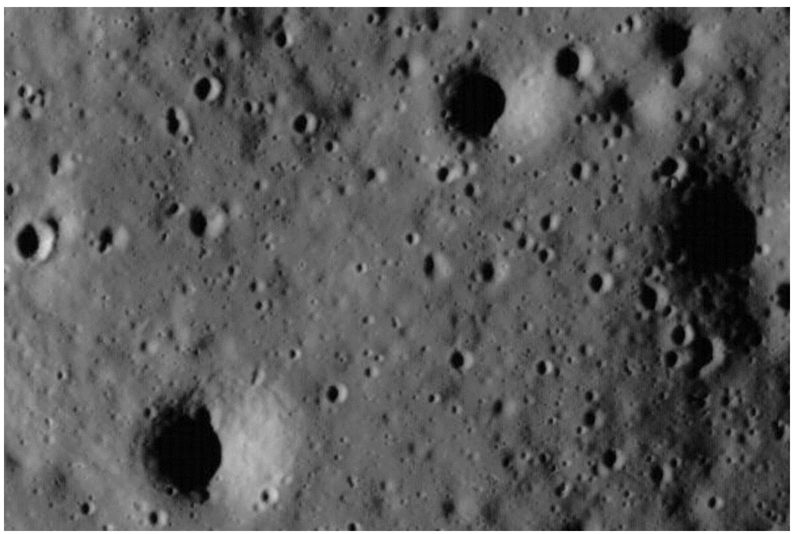
Horizon Line Detection
Robot Localization

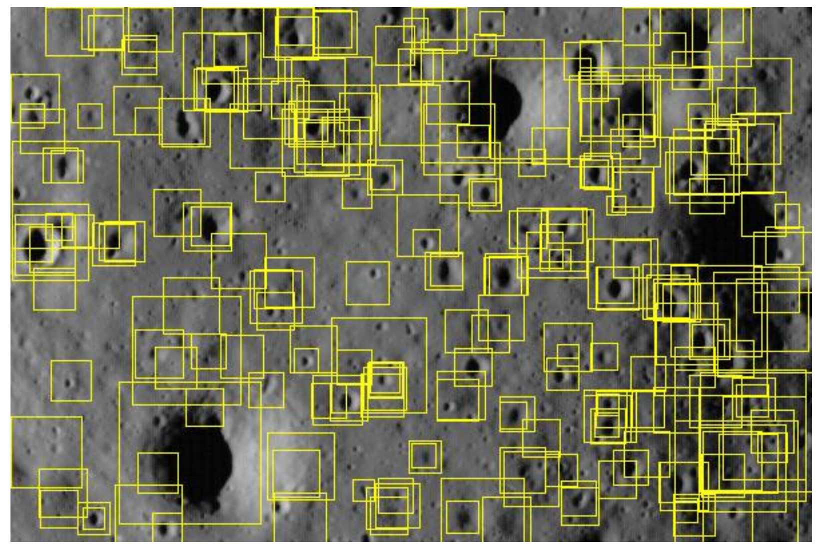


Touqeer Ahmad, George Bebis, Monica Nicolescu, Ara Nefian, and Terry Fong, "Horizon Line Detection using Supervised Learning and Edge Cues", **Computer Vision and Image Understanding (CVIU)**, vol. 191, 2020.

Related Work @ CVL

Crater Detection



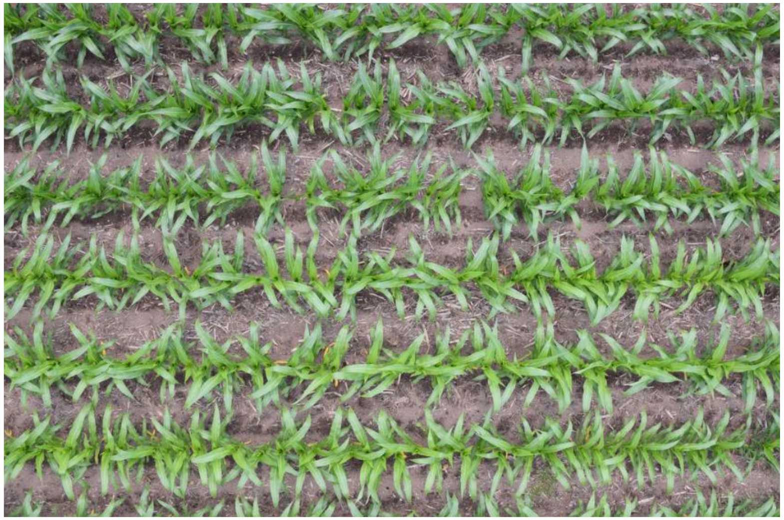


Ebrahim Emami, Touqeer Ahmad, George Bebis, Ara Nefian, and Terry Fong, "Crater Detection Using Unsupervised Algorithms and Convolutional Neural Networks", **IEEE Transactions on Geoscience and Remote Sensing**, vol. 57, no. 8, 2019.

Related Work @ CVL

Nitrogen Deficiency Detection & Quantification

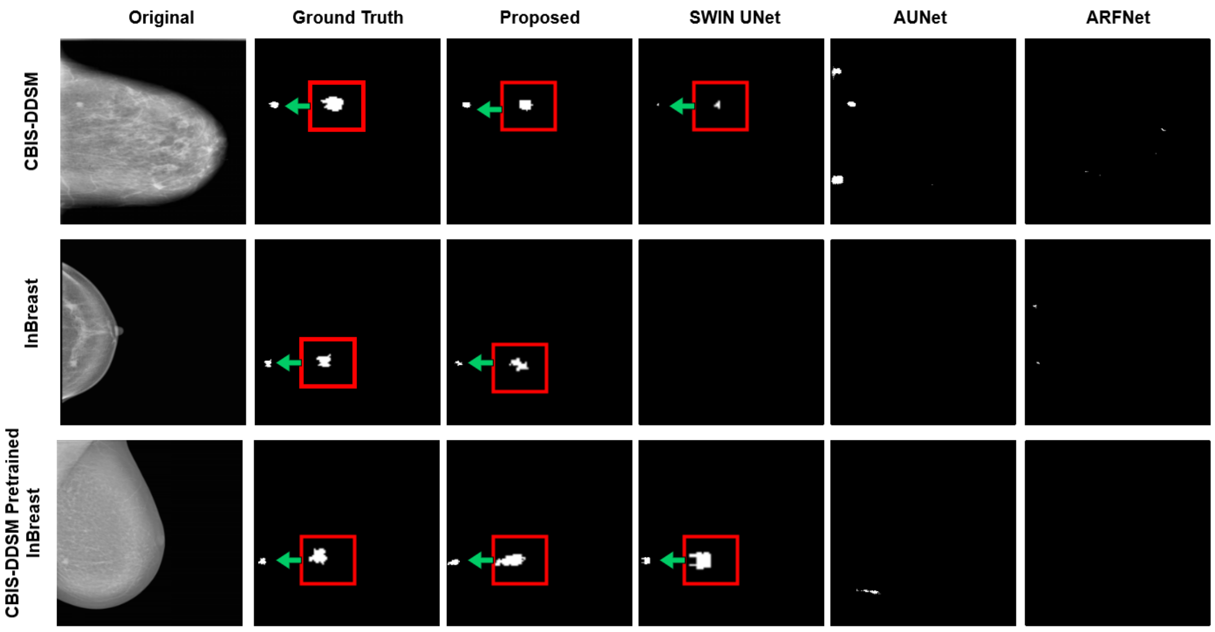
USDA NIFA



Aminul Huq, Dimitris Zermas, and George Bebis, Identification of Abnormality in Maize Plants from UAV Images Using Deep Learning Approaches", **18th International Symposium on Visual Computing (ISVC'23)**, Lake Tahoe, Nevada, USA, October 16-18, 2023.

Related Work @ CVL

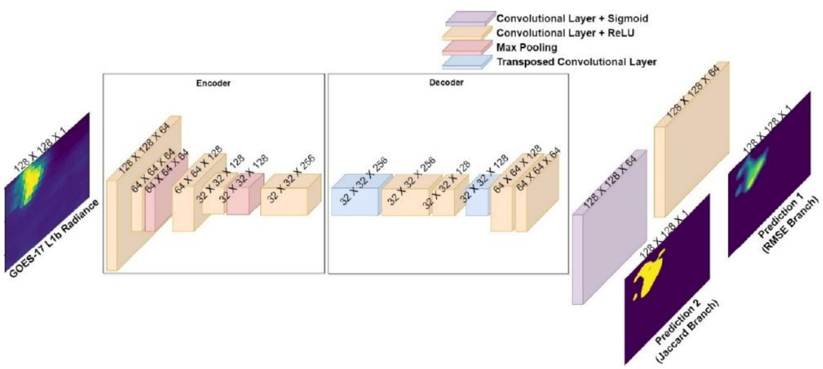
Whole Breast Micro Mass Detection and Segmentation

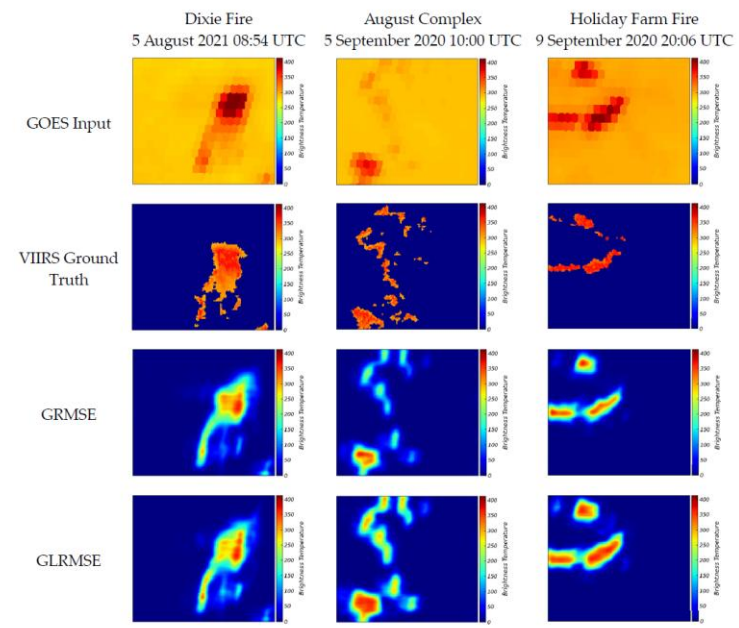


Sharif Amit Kamran, Khondker Fariha Hossain, Alireza Tavakkoli, George Bebis, Sal Baker, [SWIN-SFTNET: Spatial Feature Expansion and Aggregation Using SWIN Transformer for Whole Breast Micro-Mass Segmentation](), **IEEE International Symposium on Biomedical Imaging (ISBI 2023)**, Cartagena de Indias, Colombia, April 18-21, 2023.

Related Work @ CVL

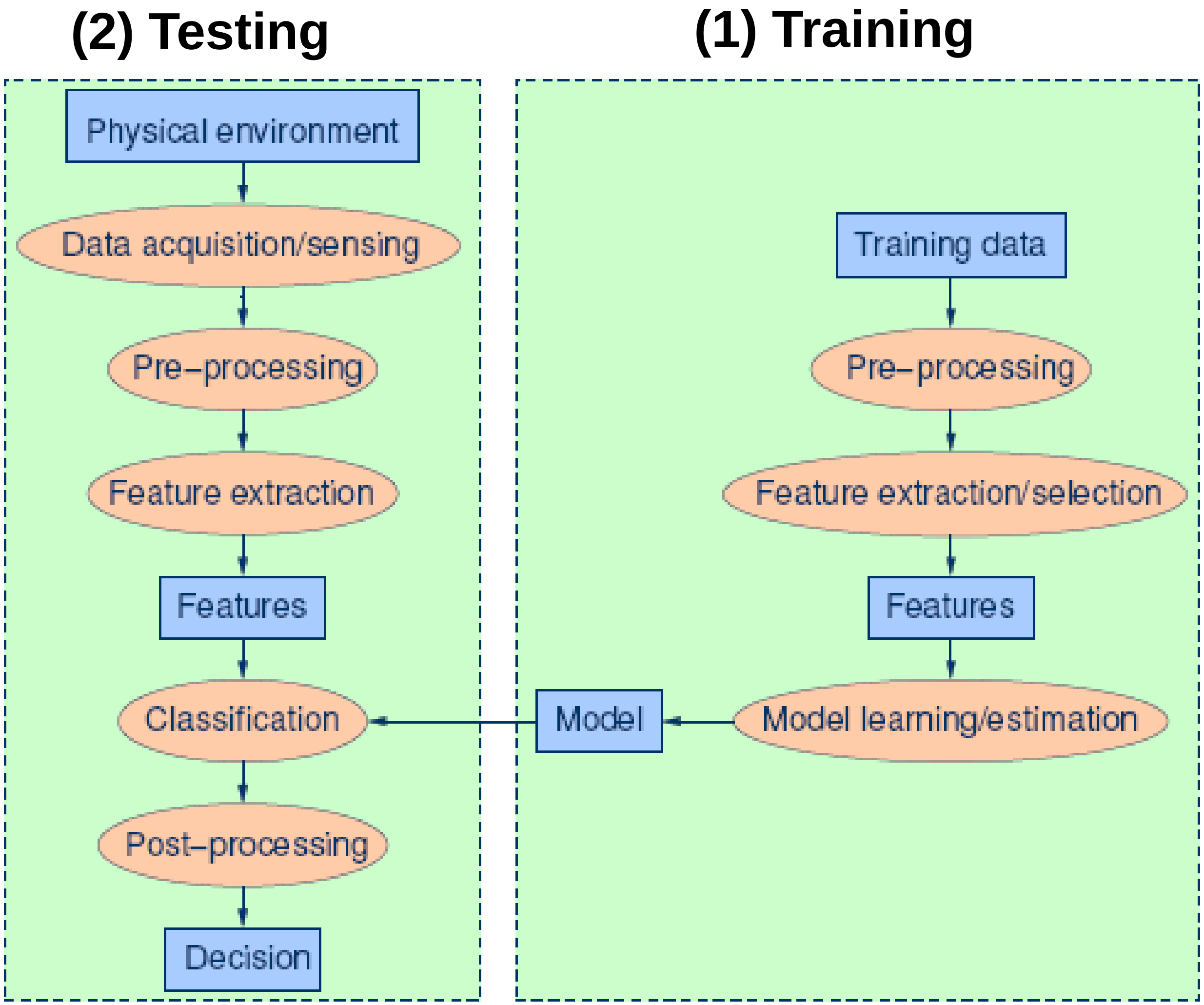
Wildfire Boundary Detection





Mukul Badhan, Kasra Shamsaei, Hamed Ebrahimian, George Bebis, Neil P. Lareau, and Eric Rowell, "Deep Learning Approach to Improve Spatial Resolution of GOES-17 Wildfire Boundaries Using VIIRS Satellite Data", **Remote Sensing**, 16(4), 2024.

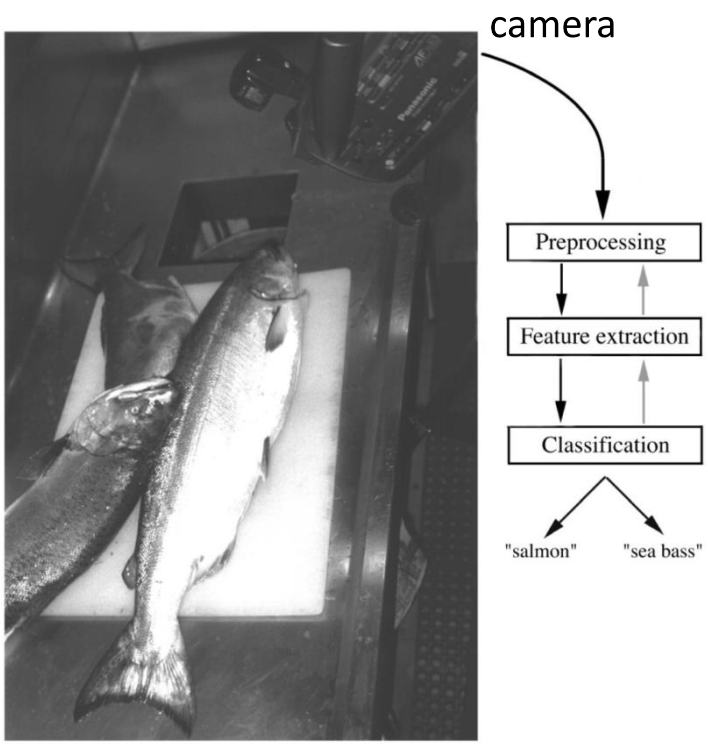
Main Phases of a PR System

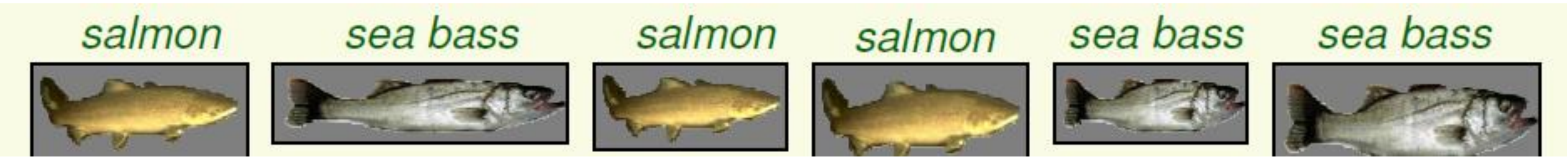


Complexity of PR – An Example

Problem: Sort incoming fish on a conveyor belt.

Two classes:
(1) sea bass
(2) salmon

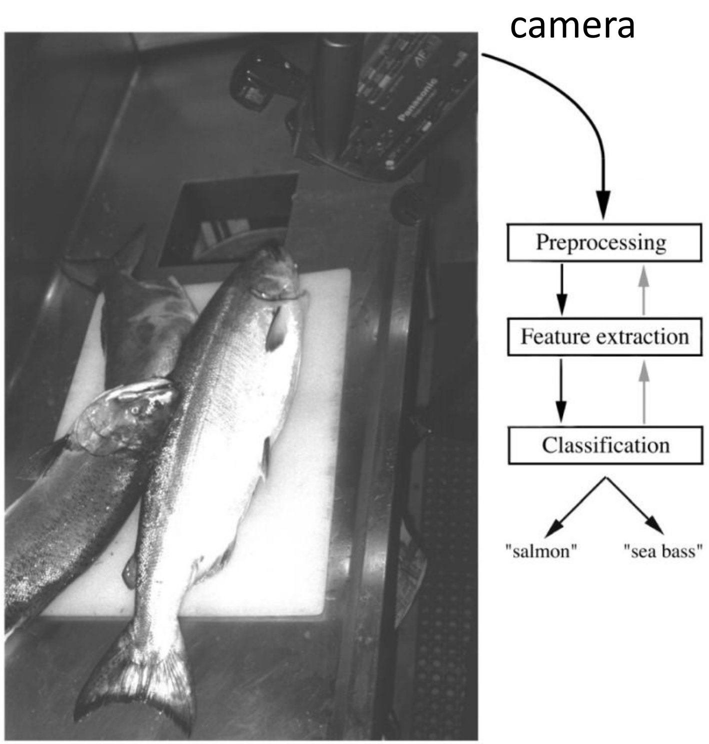




Sensors

- Select an appropriate sensor (e.g., camera, weight scale) for data collection!

- Overall performance depends on the resolution, sensitivity, and distortion of the sensor being used.



Preprocessing

A **critical** step for reliable feature extraction!

- Noise removal
- Image enhancement
- Separate touching or occluding fish

CS 476/676 – Image Processing (offered every Fall)

Training / Test data

- Need to collect a sufficiently large and representative set of examples for **training** and **testing** the system!

Training Set ?

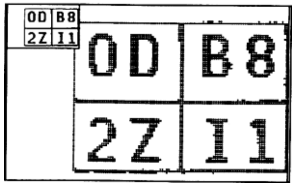
Test Set ?

Data Variability

- Intra-class variability

The letter "T" in different typefaces

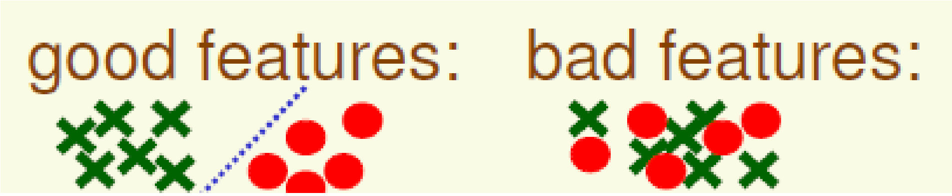
- Inter-class variability

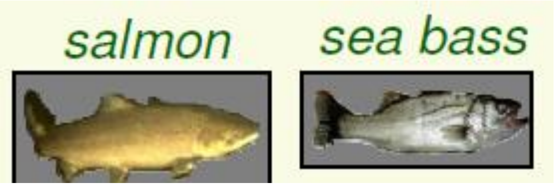


Letters/Numbers that look similar

Feature Extraction

- How should we choose a good set of features?
  - Discriminative features





  - Invariant features (e.g., invariant to geometric transformations such as translation, rotation and scale)

- Are there ways to automatically learn which features are best ?

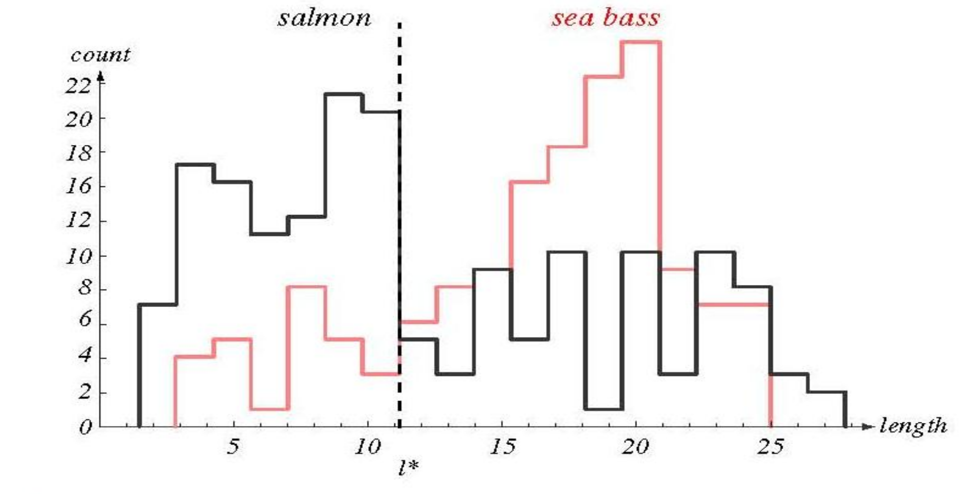
Feature Extraction - Example

- Assume that sea bass is generally longer than salmon.
  - Use length as a feature!
  - Decide between sea bass and salmon by applying a threshold on length.

- How should we choose the threshold?

Feature Extraction - Example (cont'd)

Histogram of "length"



threshold l*

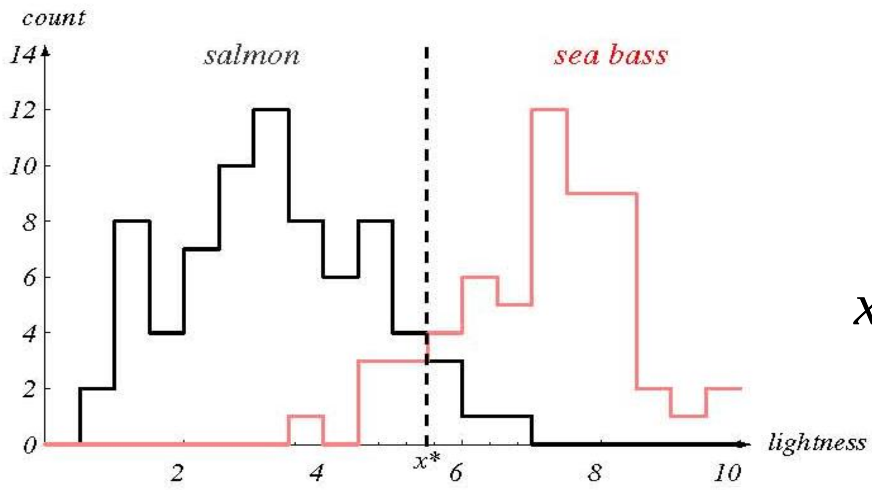
$x_1 : length$

- Even though sea bass is longer than salmon on average, this is **not** always the case (i.e., distributions overlap).

Feature Extraction - Example (cont'd)

- Let's consider a different feature: lightness

Histogram of "lightness"



threshold x^*

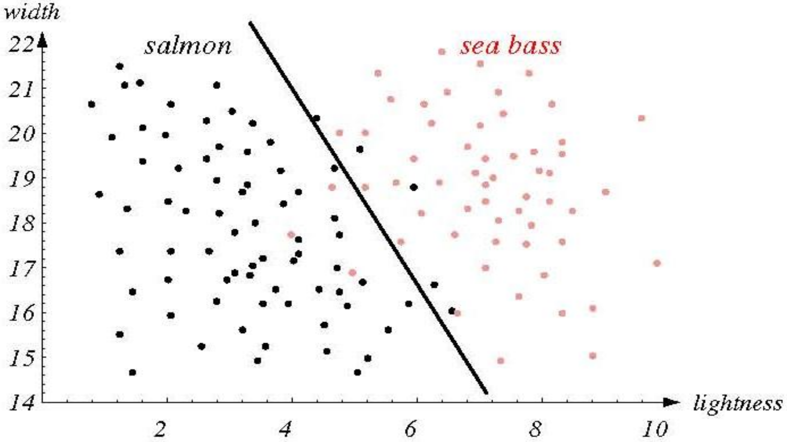
$x_1 : lightness$

- Easier to choose a threshold but we still cannot make a perfect decision.

Multiple Features

- Use additional features!
  - Single features might not yield good performance.
  - Combinations of features might yield better performance.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \begin{matrix} x_1 : lightness \\ x_2 : width \end{matrix}$$

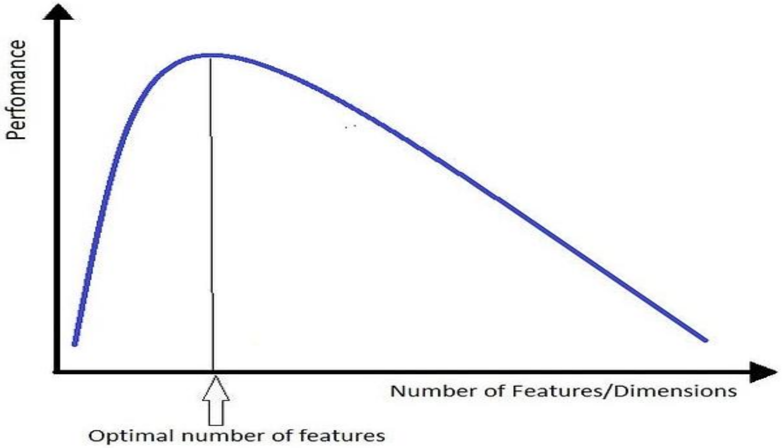


Multiple Features (cont'd)

- Does adding more features always help?
  – It might be difficult and computationally expensive to extract more features.
  – Correlated and irrelevant features will not improve performance (i.e., redundancy).
  – Adding too many features can, paradoxically, lead to a worsening of performance – this is known as the "curse" of dimensionality!

Curse of Dimensionality

- Typically, this is what happens when we keep adding more and more features!



- It turns out that the more features we add, the more data we need for training!

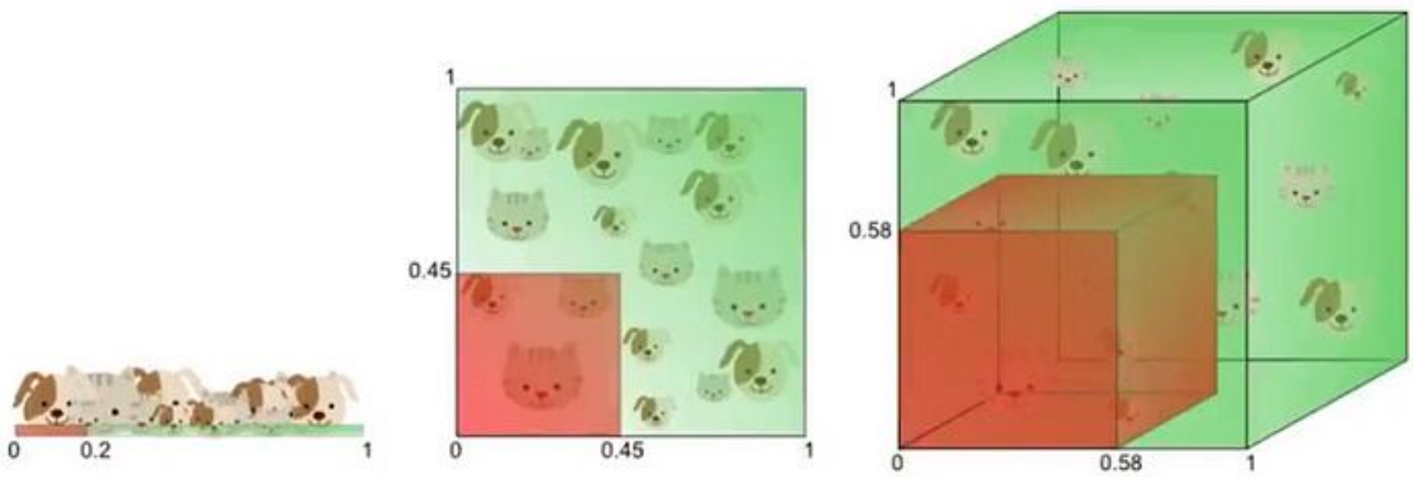
- <u>Here is the bad news</u>: the number of training data required grows exponentially with the number of features!

Let's see why!

Curse of Dimensionality (cont'd)

- Suppose we apply uniform sampling to collect the training data for each class!

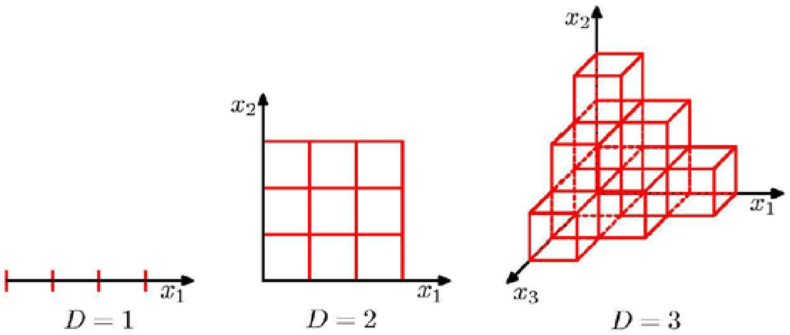
- The amount of data needed to cover some range of features (e.g., 20% of the space), grows exponentially with the number of dimensions.



D=1 **D**=2 **D**=3

Curse of Dimensionality (cont'd)

- Let's divide each feature dimension uniformly into **M** cells.



- Sample at least one data point from each cell.
- What is total number of cells? $\mathbf{M^D}$ (***D:*** # of features).
- Therefore, the number of training data grows exponentially with ***D***!

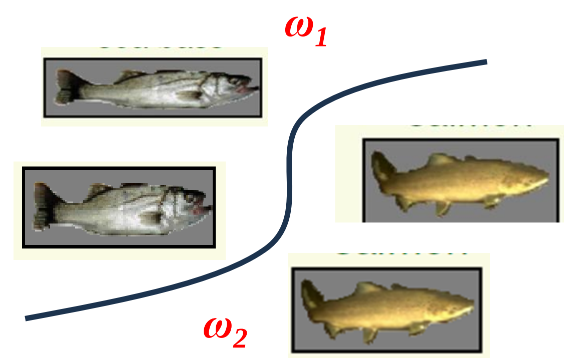
Missing Features

- Certain features might be missing (e.g., due to occlusion).
- How should we train the classifier with missing features?
- How should the classifier make the best decision with missing features?

Classification Models

***x**: feature vector (pattern)*

***ω**$_i$: class label*



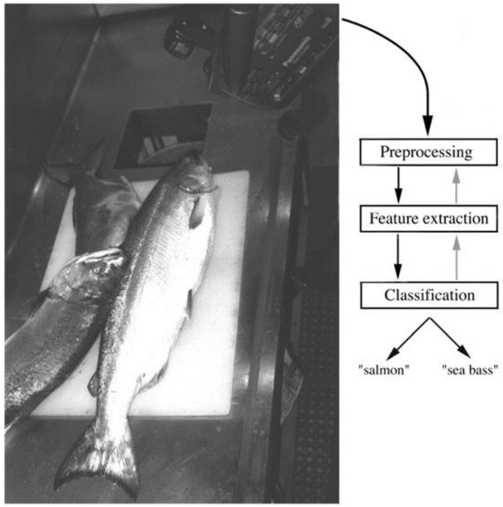
pdf: probability density function

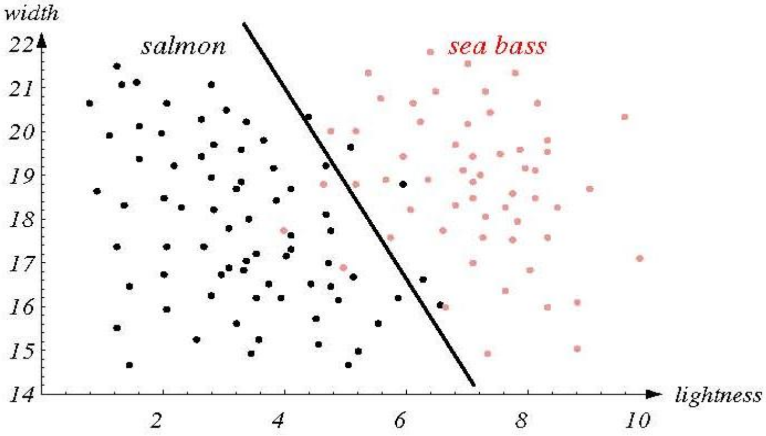
- ***Generative***
  - Model and estimate the **joint** pdf **p(x, ω$_i$)** for each class.
    - e.g., use a **Gaussian** pdf for each class
  - **Decision boundary** can be inferred from the joint pdfs.
  - Make predictions by using the **Bayes rule** to calculate **P(ω$_i$ / x)**
  - Pick the most likely class label ***ω**$_i$*

- ***Discriminative***
  - Model and estimate the **decision boundary** explicitly.
  - Calculate **P(ω$_i$ / x)** by "learning" a direct mapping from **x** to ***ω**$_i$*
  - Pick the most likely class label ***ω**$_i$*

Decision Boundary

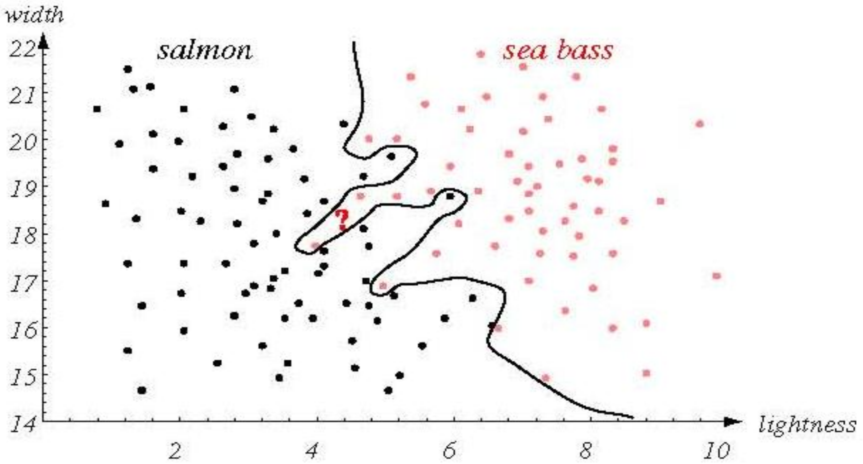
How should we find an optimal decision boundary?



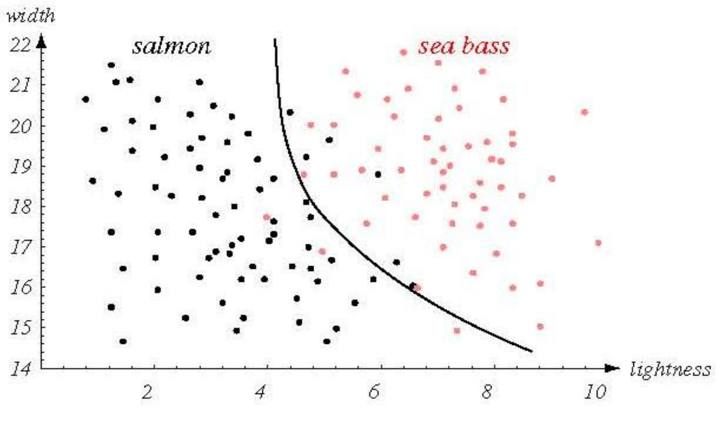


Decision Boundary (cont'd)

- We can get perfect classification results on the training data by choosing a complex model.
- Should we prefer a complex model or a simpler one?



complex model

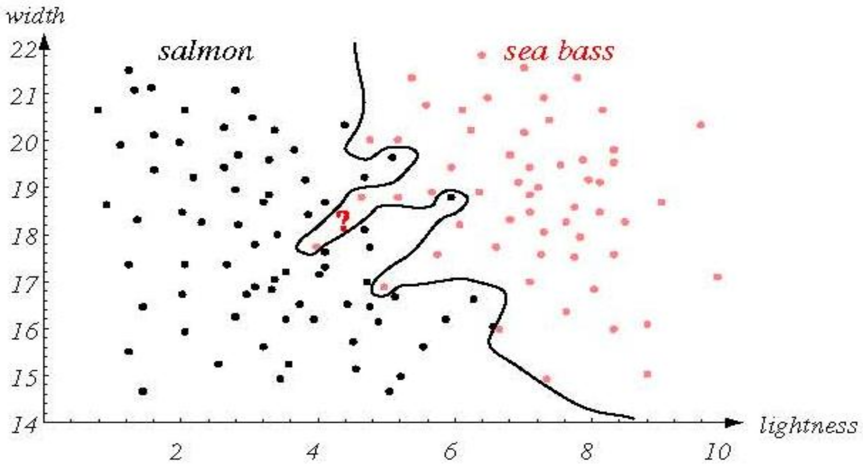


simpler model

Note: the complexity of a model depends on the number of parameters required to fully specify the model.

Overfitting

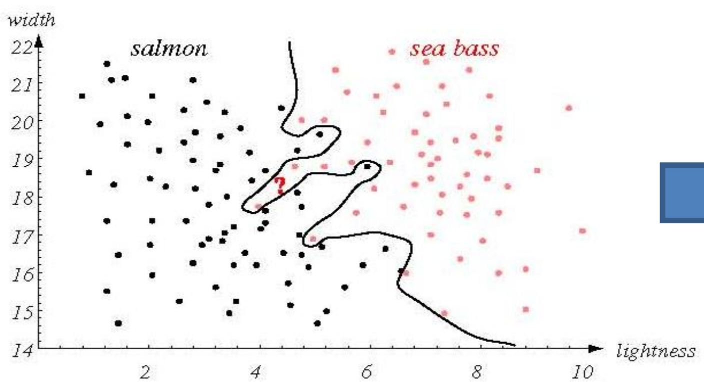
- Complex models are tuned to the training data, rather than to the characteristics of the true model (i.e., memorization or overfitting).
- Overfitting leads to poor generalization!



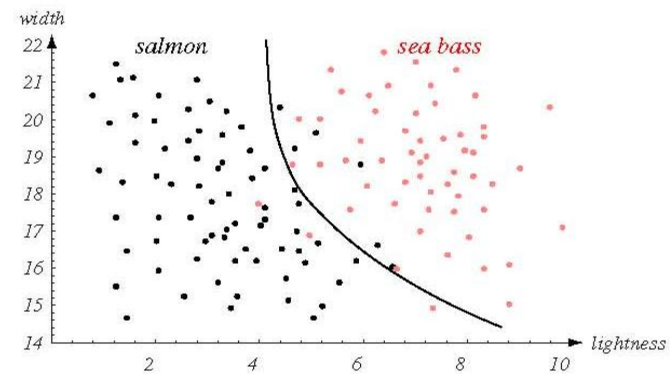
complex model

Generalization

- Generalization is defined as the ability of a classifier to correctly classify novel patterns (i.e., **not** in the training set).
- How could generalization performance be improved?
  - Using more training data (i.e., leads to better model estimation).
  - Using simpler models (i.e., involves fewer model parameters).



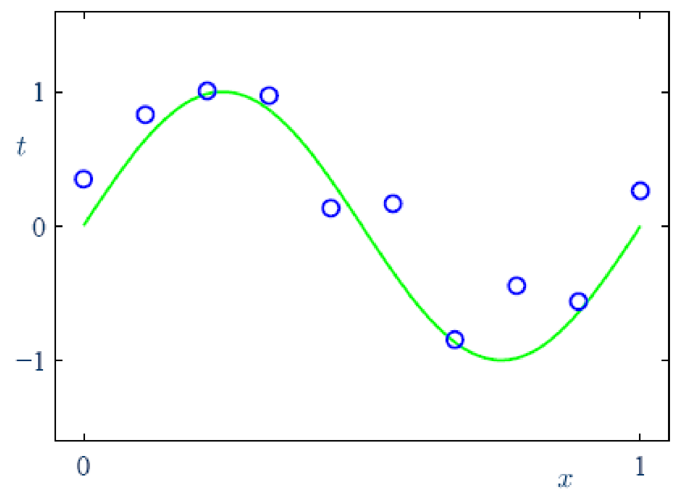
complex model



simpler model

Understanding model complexity: function approximation

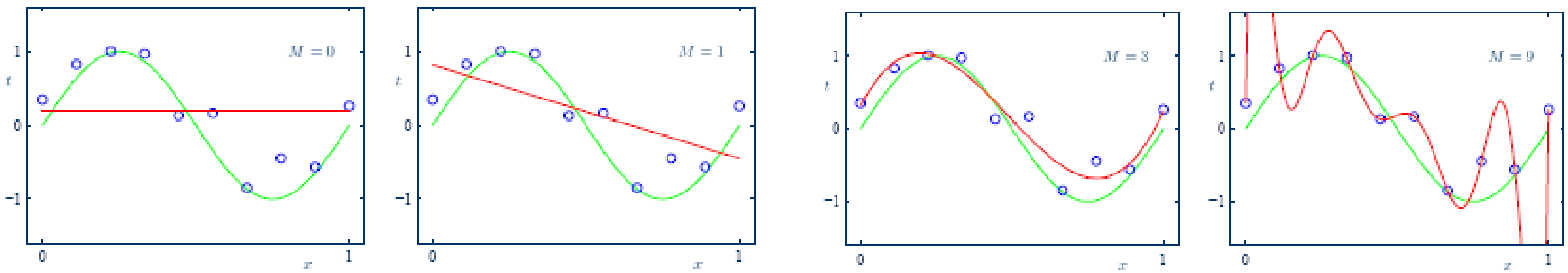
- Approximate a function from a set of samples:
  - Green curve is the **true** function (i.e., model).
  - Ten sample points are shown by the blue circles.
  - Samples typically contain some "noise" (i.e., do not lie on the "green" curve).



Understanding model complexity: function approximation (cont'd)

Polynomial curve fitting: polynomials having various orders (i.e., complexity/parameters), shown as red curves, fitted to the samples.

overfitting



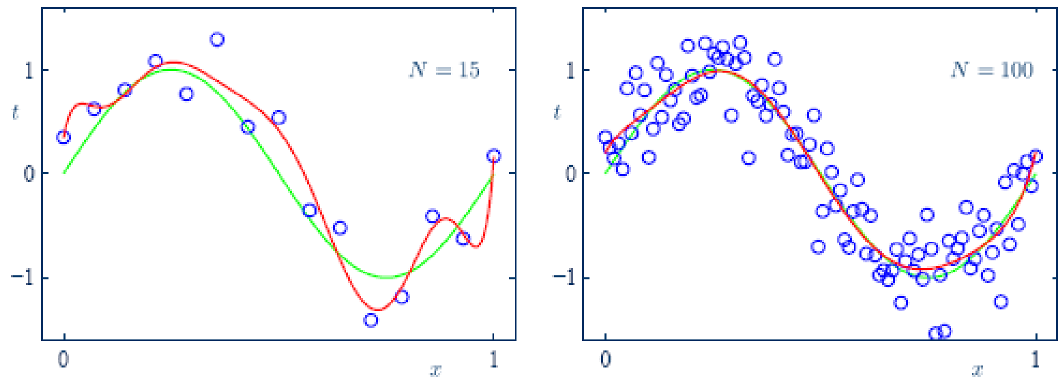
(a) 0'th order polynomial (b) 1'st order polynomial (c) 3'rd order polynomial (d) 9'th order polynomial

In the absence of enough data, simpler models are preferred over complex models.

Understanding model complexity: function approximation (cont'd)

- More data can improve model estimation as shown below!

9^{th} order polynomials fitted to **15** and **100** samples.



(a) 15 sample points (b) 100 sample points

Cost of miss-classifications

- There are two possible classification errors in the fish classification example:

  (1) Deciding the fish was a sea bass when it was a salmon.

  (2) Deciding the fish was a salmon when it was a sea bass.

- Are both errors equally important in practice?

Cost of miss-classifications (cont'd)

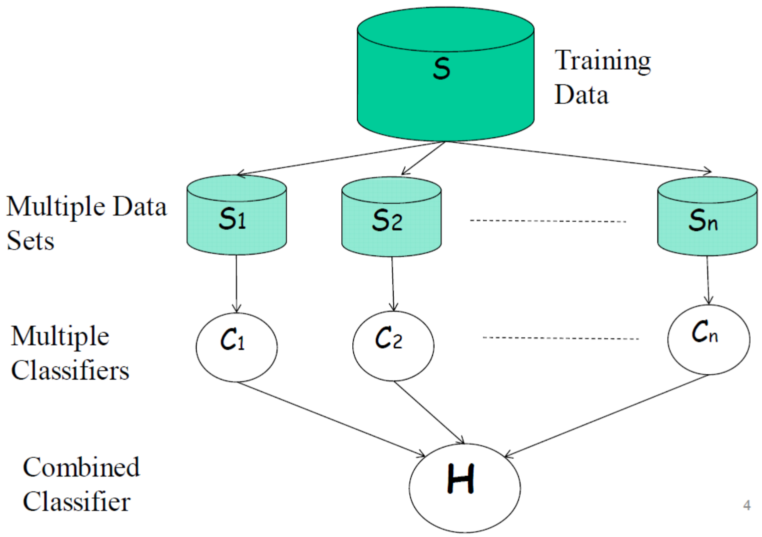
- Let us assume that:
  - Customers who buy sea bass will object vigorously if they see salmon in their cans.
  - Customers who buy salmon will not be unhappy if they occasionally see some expensive sea bass in their cans.

- How does this knowledge affect classification design?

Improve Classification Performance using Ensembles of Classifiers

- Performance can be improved using a "pool" of classifiers.

- How should we build and combine different classifiers ?



Improve Classification Performance through Post-processing

- Consider the problem of character recognition.

How m ch info mation are y u mi sing?

- Exploit context to improve classification accuracy!

Computational Complexity

- How does an algorithm scale with:
    - Number of features
    - Number of training data
    - Number of classes

- Need to consider tradeoffs between computational complexity and performance.

Would it be possible to build a "general purpose" PR system?

- Very difficult to design a system that is capable of performing a variety of classification tasks.
  - Different problems require different features.
  - Different features might yield different solutions.
  - Different tradeoffs exist for different problems.

Quiz #1

- When: Monday, Feb 3rd
  - Closed book/notes, 4 questions, 15 minutes
- What: Introduction to Pattern Recognition
- Questions will test your understanding of the **main** concepts related to the above material.