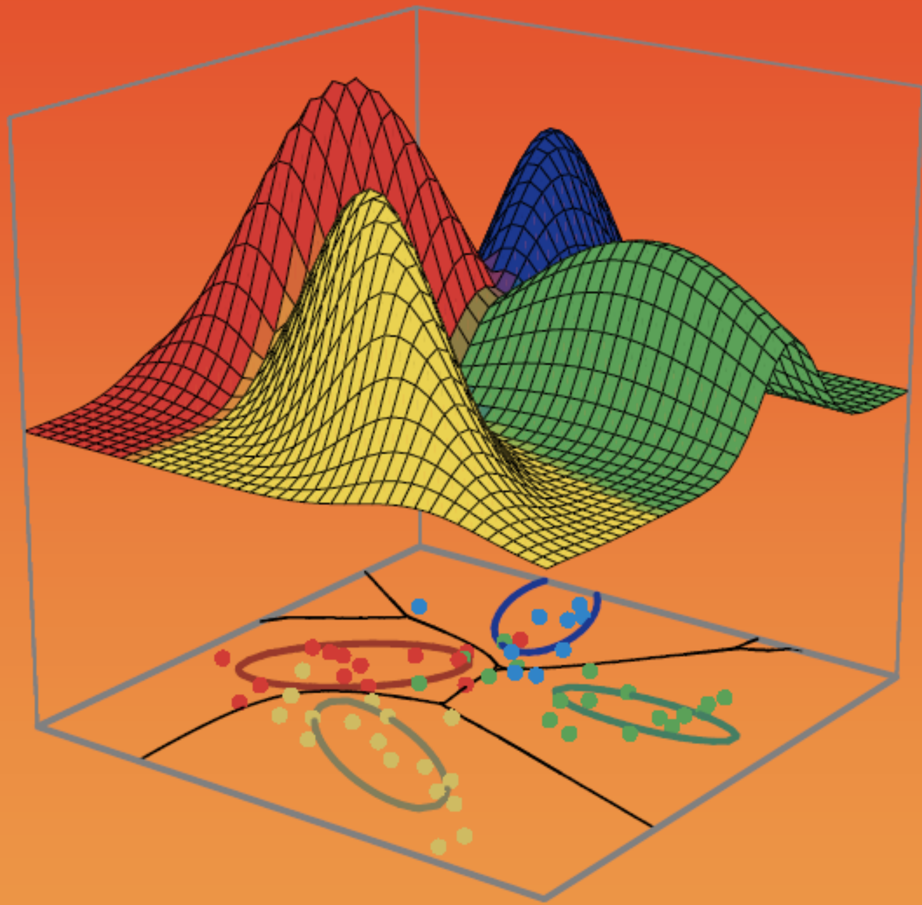
Pattern Recognition
ECSE 4410/6410 CAPA
Spring 2022

<u>Supportive Material</u>
Part I

Course Instructor - Thirimachos Bourlai

January to May 2022

Please Check the 2022 Syllabus



Pattern Classification

A lot of material in these slides was taken from **Pattern Classification (2nd ed)** by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000 with the permission of the authors and the publisher

Some slides are taken from CS 559: Machine Learning Fundamentals and Applications - Philippos Mordohai, Webpage: www.cs.stevens.edu/~mordohai

What is Pattern Recognition

- Informally
  – Recognize patterns in data (1D, 2D etc. signals)
- More formally
  – **Assign** an **object** or an **event** **<u>to</u>** one of the several pre-specified **<u>categories</u>**
    > A category is usually called a **class**

- **Duda et al. Book – Definition** : "Pattern recognition — the act of taking in raw data and taking an action based on the ***category*** of the pattern"

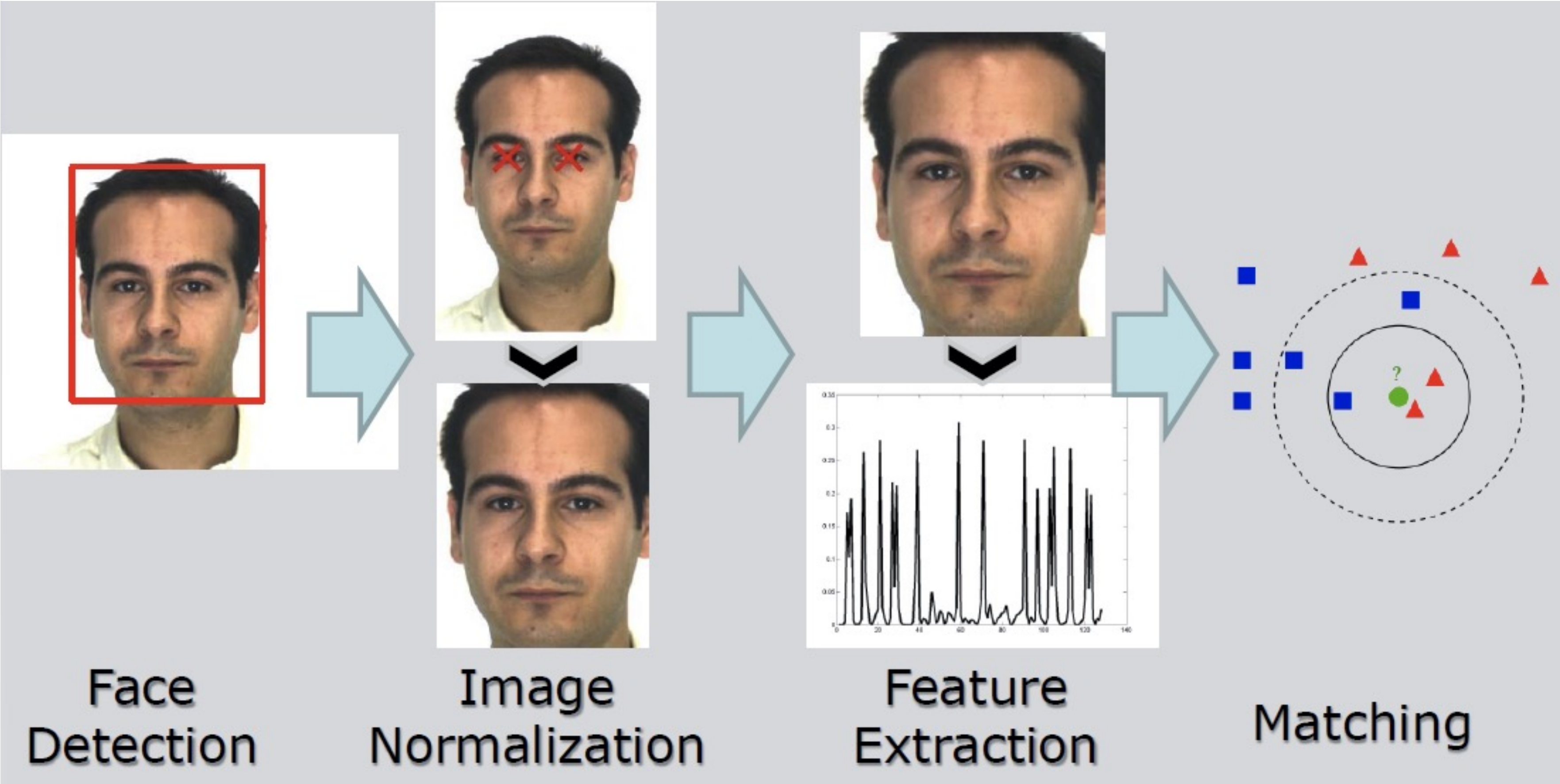
Acts of Pattern Recognition

- **Understand** spoken words
- **Read** handwritten characters
- **Identify** our car keys in our pocket by feel
- **Decide** whether an apple is ripe by its smell
- **Identify** individuals (e.g. Face Recognition)

Character Recognition

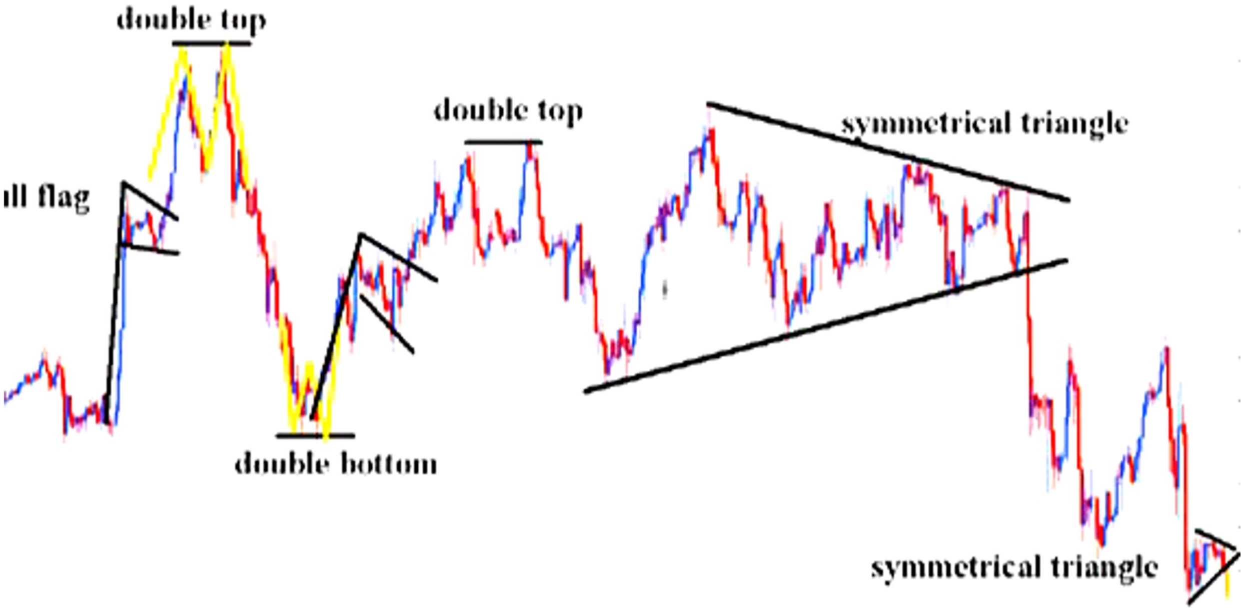
- In this case the classes are all characters in the alphabet, digits etc.

How Automated FR Works [4]



[4] A. K. Jain, “Automatic Face Recognition: State of the Art”, Korea University, Seoul, Sept. 15, 2010.

GBP/JPY Patterns!



Machine Learning Research

- Speech recognition
- Natural language processing
- Computer vision
- Medical outcomes analysis
- Robot control
- Computational biology
- Sensor networks

DARPA – Autonomous Vehicles

DARPA's BigDog robots, seen here trotting around in the shadow of an MV-22 Osprey while given commands via remote control at Marine Corps Air Station New River, N.C., June 26, 2006, **could carry extra gear to free soldiers and Marines of the burden of extra weight**.

DARPA – Autonomous Vehicles

Eric Krotkov, DARPA's UGV program manager in 1997-1998, helped refocus the concept of autonomous ground robots by bringing to the military effort a decade of experience working on planetary rovers for NASA.

The Crusher unmanned ground vehicle (UGV) is being developed under the DARPA/Army UGCV-Perception for Off-Road Robotics Integration (UPI) program. Crusher is a highly mobile vehicle designed from the outset to be unmanned. It is being equipped with state-of-the-art perception capabilities, and will be used to **validate the key technologies** necessary for an unmanned ground vehicle to **perform military missions autonomously**.

Discussion

- Any examples that you can provide?

Every-Day-Life Applications

- Loan applications
- Recommendation systems
  - Amazon, Netflix
- Targeted advertising
  - Countless examples…

<u>Chapter 1</u>

Introduction to Pattern Recognition *(Sections 1.1-1.6)*

- Machine Perception
- An Example
- Pattern Recognition Systems
- The Design Cycle
- Learning and Adaptation
- Conclusion

Machine Perception

- Build a machine that can recognize patterns:
  - Speech Recognition
  - Computer Vision: Object recognition, Face detection
  - Fingerprint Identification
  - OCR (Optical Character Recognition)
  - DNA sequence identification

Example

- Fish packing plant wants to **automate the process of sorting incoming fish** on a conveyor belt according to species
- Try to separate sea bass from salmon using **optical sensing**
- **Physical differences** between the two types of fish:
  - Length, lightness, width, number and shape of fins, mouth position

Suggestion: *"To use these features to explore for use in our classifier"*

How about noise or variations in the images?

[Atlantic Salmon (*Salmo salar*)]

[Brook Trout (*Salvelinus fontinalis*)]

[Brown Trout (*Salmo trutta*)]

[Lake Trout (*Salvelinus namaycush*)]

[Rainbow Trout or Steelhead (*Oncorhynchus mykiss*)]

[Chinook Salmon (*Oncorhynchus tshawytscha*)]

Model

- **There are differences** between sea bass and salmon
  - Thus, **there are different models** (different descriptions), which are typically mathematical in form

- **Pattern Classification Goal**:
  1) **Hypothesize** the class of these models
  2) **Process** the sensed data to eliminate noise
  3) For any sensed pattern **choose** the **model** that corresponds best

Discussion

- How would you build a PR system?

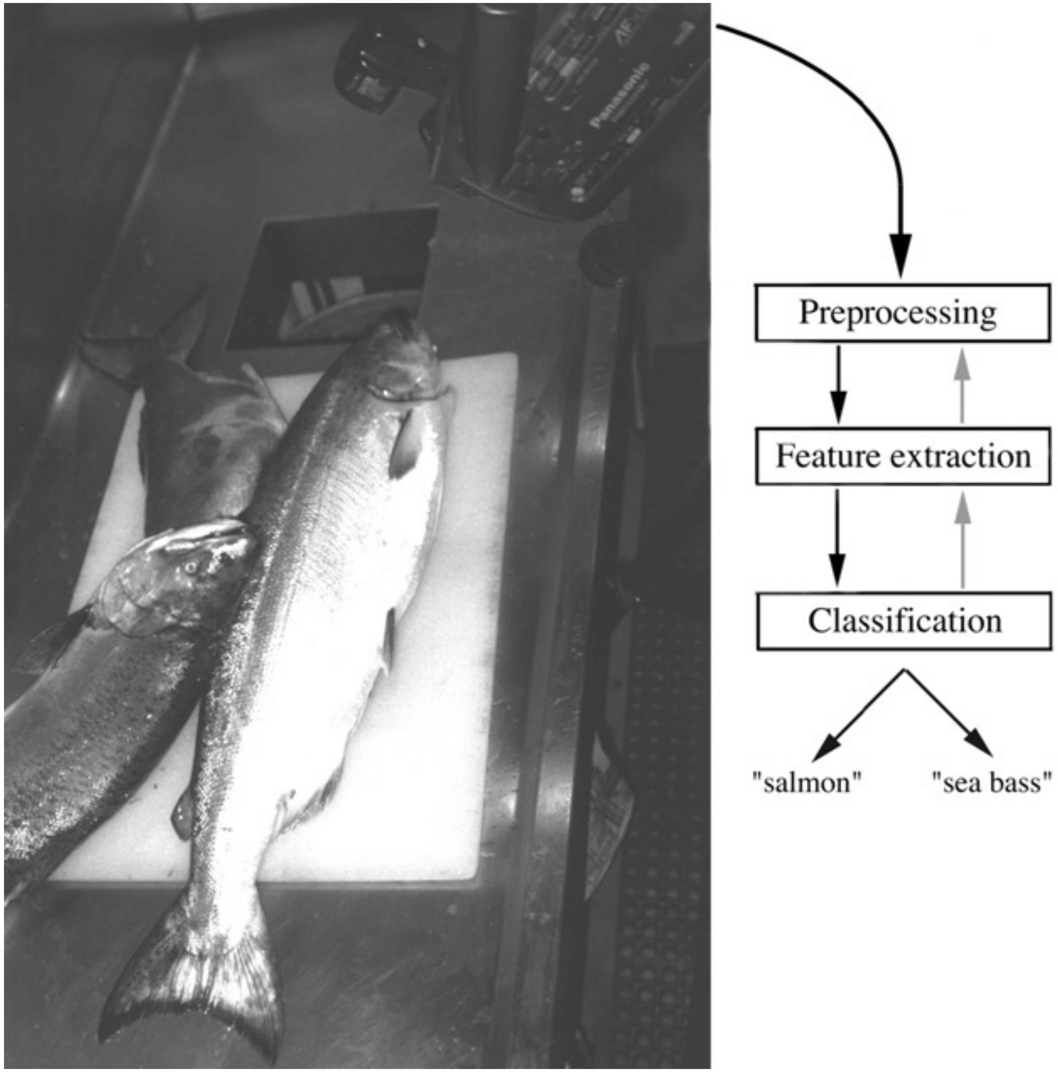


Figure: **System Overview** - The objects to be classified are first sensed by a transducer (camera). Then, images are preprocessed, features are extracted and, finally, classification is performed (here either “salmon” or “sea bass”)

Training

- Set up a camera and take some sample images
  - Label these images by hand

- Extract features
  - Length
  - Lightness
  - Width
  - Number and shape of fins
  - Position of the mouth, etc…

  - Test whether this set of features is useful for a classifier

Preprocessing

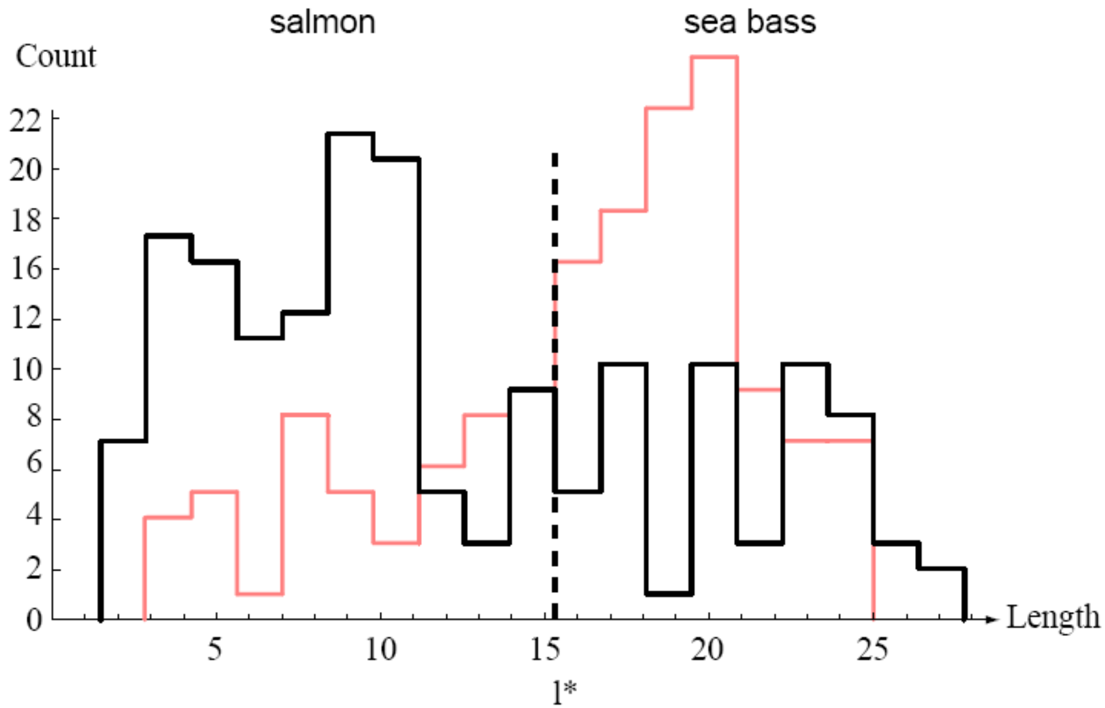
- Use a segmentation operation to isolate fishes from one another and from the background

- Information from a single fish is sent to a feature extractor whose purpose is to reduce the data by measuring certain quantities

- The features are passed to a classifier

Discussion

- How can you interpret the graph and the outcome of this classifier?



Design the Classifier
- Using only 1 feature, i.e. the "fish length"

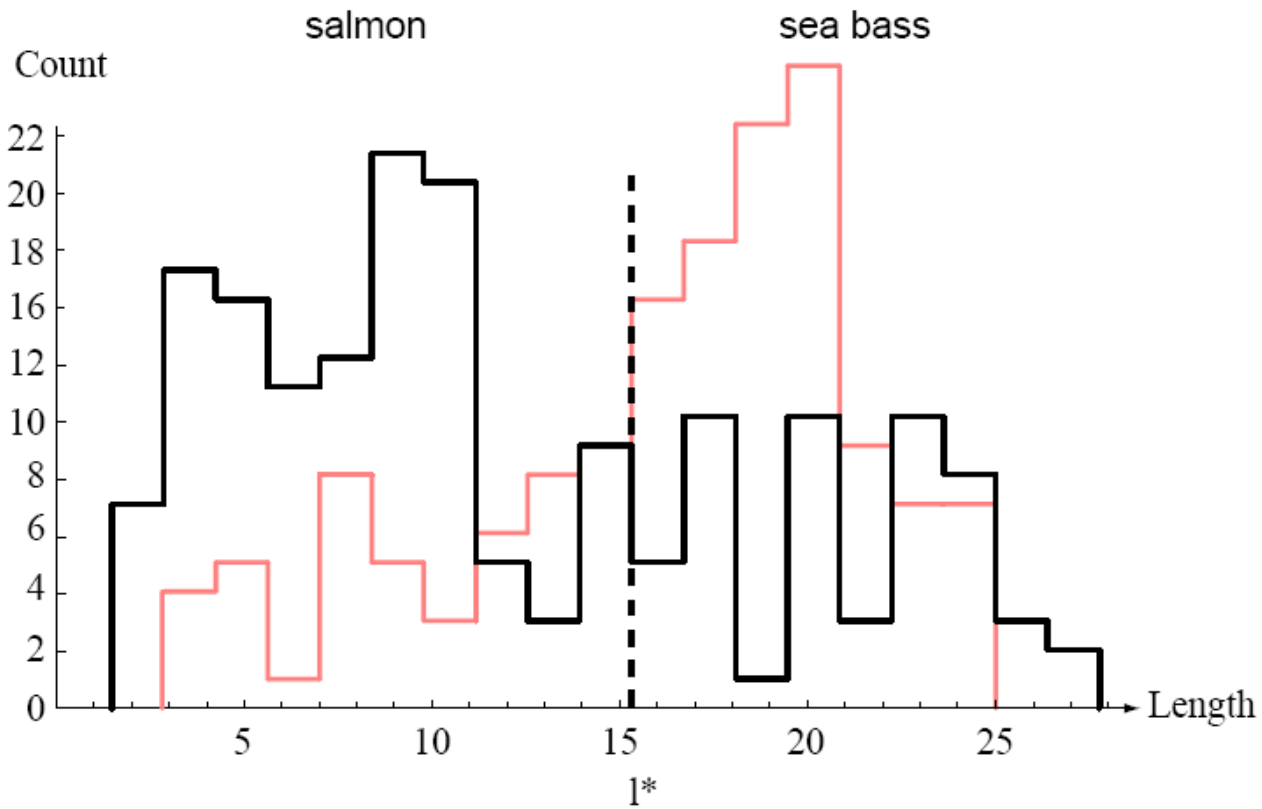


Figure: Histograms for the length feature for the two categories. No single threshold value l∗ (decision boundary) will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value l∗ marked will lead to the smallest number of errors, on average.

Preliminary Results

- Length is a poor feature alone!
  - About 20% misclassification rate at best threshold choice

Not Good!

Discussion

- Can we use any other features when we design the classifier?
- Any ideas?

Discussion

- Can we eliminate variations in illumination?

Design the Classifier
- Using Fish Length vs. Lightness

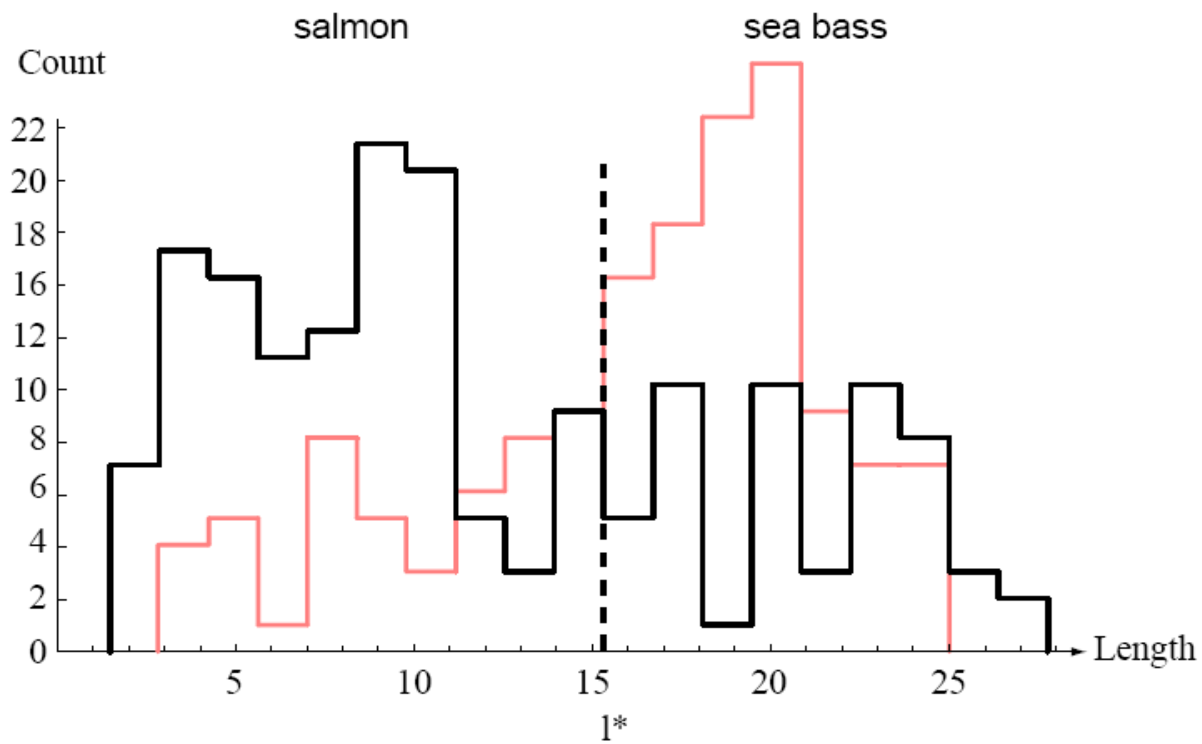


Figure : Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors, on average.

The role of "COST" in Classification

Question: were the consequences of our actions <u>equally costly</u>?

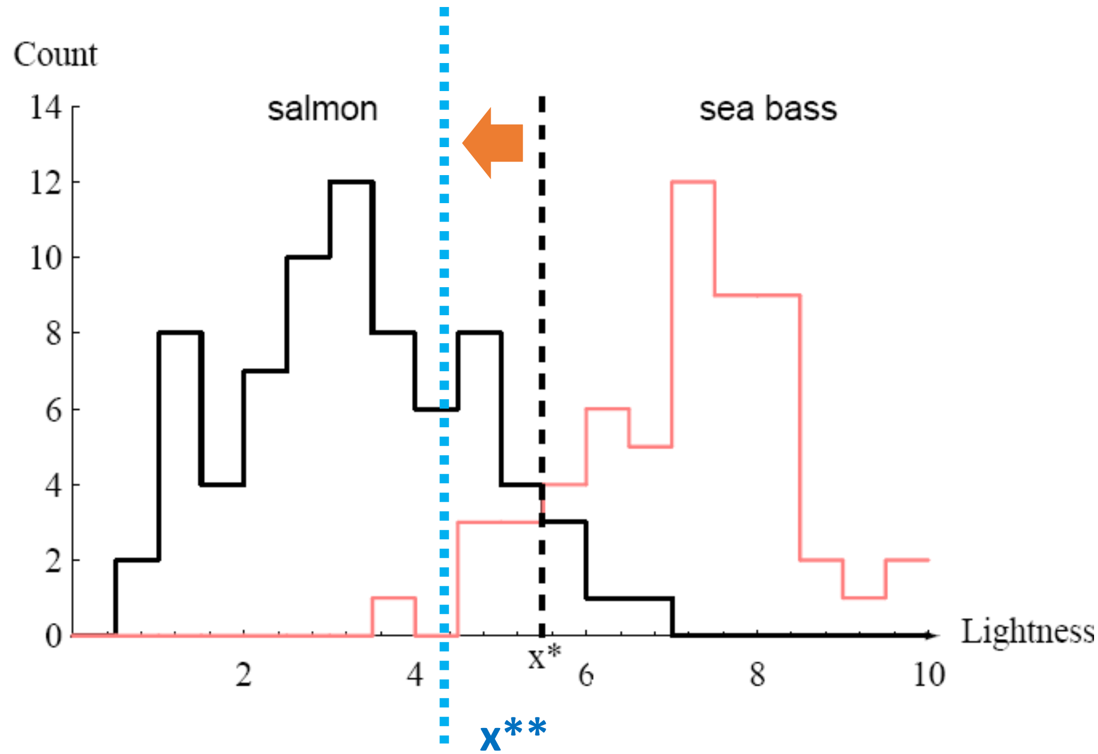
1. Decide FISH class (salmon) — Actual FISH class (sea bass)
2. Decide FISH class (sea bass) — Actual FISH class (salmon)

Same Cost

THIS IS NOT ALWAYS RIGHT!

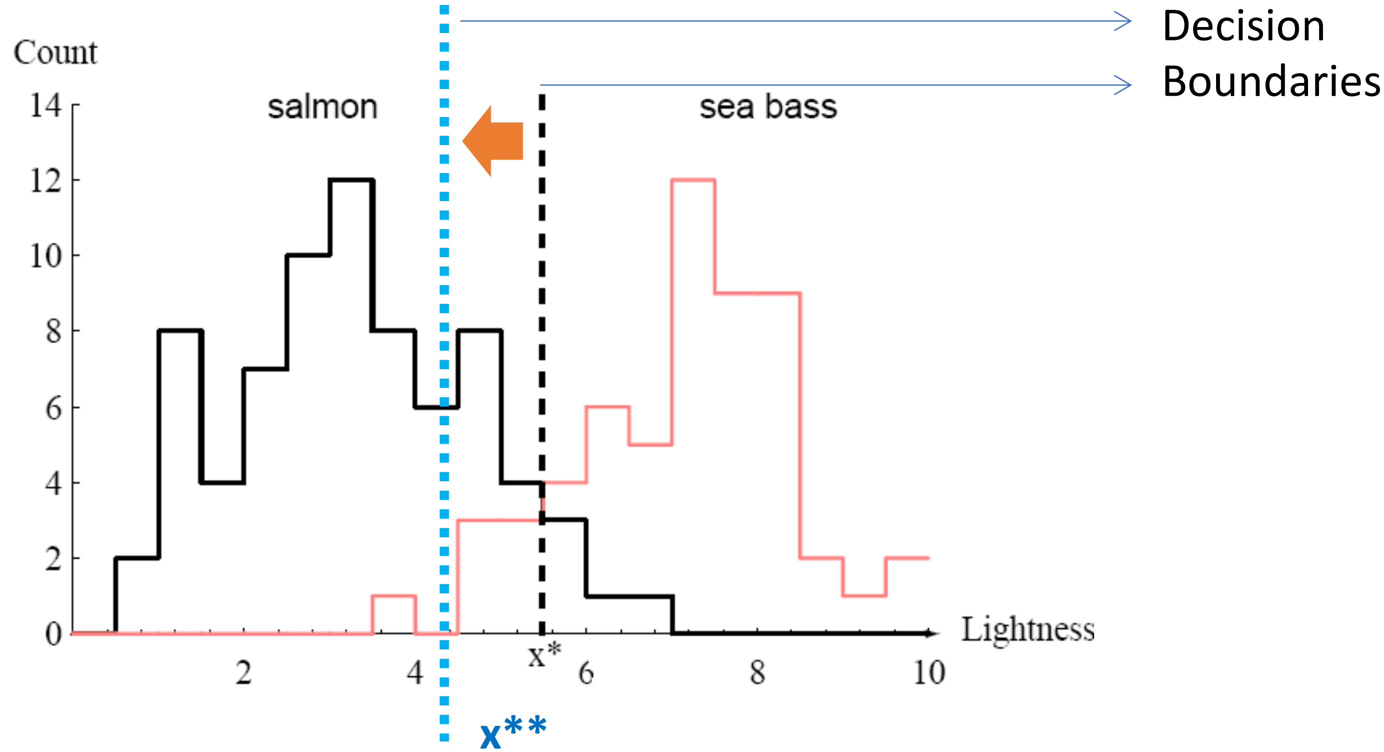
Labeling fish as Salmon (>expensive), i.e. case 1 above, can be <u>more acceptable</u> (less cost for the receiver of the fish) than case 2.

Changing the "Error COST"



- Customers <u>object</u> in having sea bass with their salmon
  - ⇒ Such Error has More Cost
  - ⇒ We set: x* → x**
  - ⇒ Customers **will** get <u>Less Sea Bass with their Samlon</u>

In PR Terms…



– Move the decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of misclassifications)

Costly!

Decision Theory

ONE OF

THE MOST IMPORTANT SUBFIELDS OF

PATTERN RECOGNITION

New Classifier

- Adopt the lightness and add the width of the fish

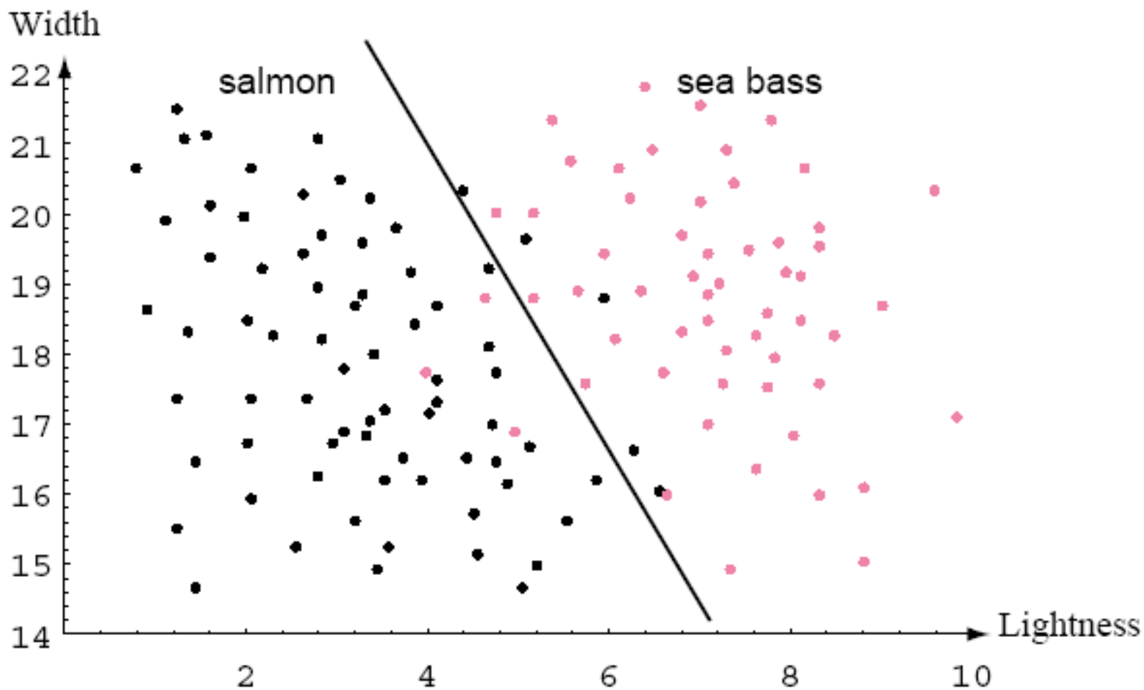
$$\text{Fish} \longrightarrow x^T = [x_1, x_2]$$

Lightness Width

Feature vector x in a <u>two-dimensional feature space</u>

New Classifier

Suppose that we measure the feature vectors for our samples and obtain the scattering of points



This plot suggests the <u>following rule</u> for <u>separating the fish</u>:
"Classify the fish as sea bass if its feature vector falls above the *decision boundary shown, and as salmon otherwise*"

Important Points

- **Other features?**
    - Width of the fish
    - The vertex angle of the dorsal fin (shape parameter)
    - The placement of the eyes (as expressed as a proportion of the mouth-to-tail distance)

- **How do we know beforehand which of these features will work best?**

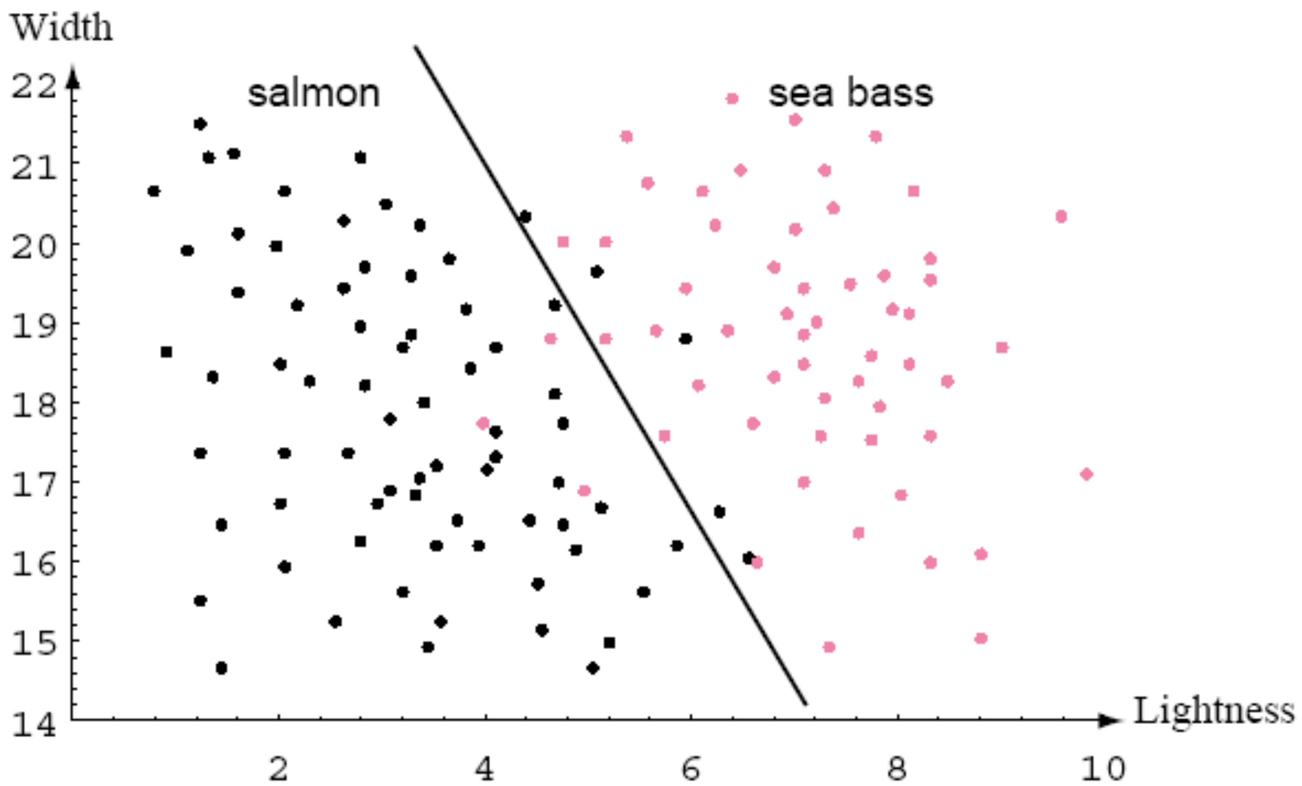
- **Redundant features?**

e.g. if the *eye color* of all fish **correlated perfectly with** *width*, then classification performance **need not be improved** if we also include eye color as a feature

- **Computational cost in attaining more features?**

    * If it is of no concern, can we have more *features and how many?*

Fast Results > Few Features > Easy Boundary

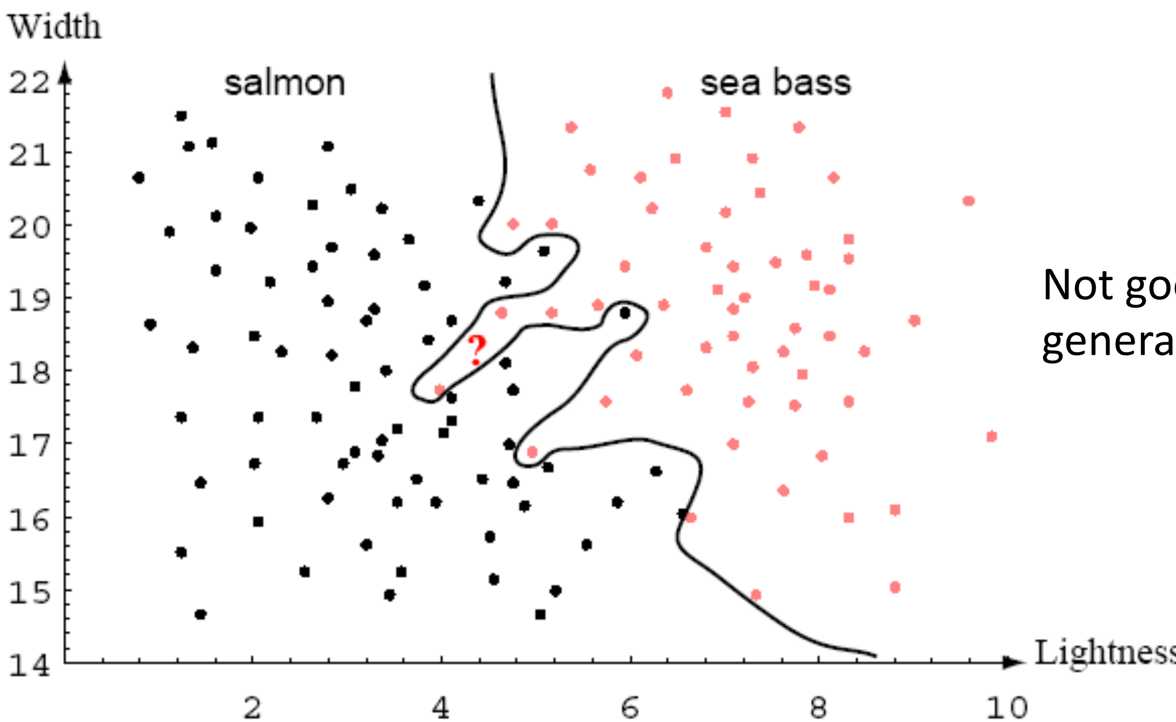


Optimal Performance?

- We may add other features that are not correlated with the ones we already have

- Intuitively, the best decision boundary should be the one which provides an optimal performance such as in the following figure:

Generalization

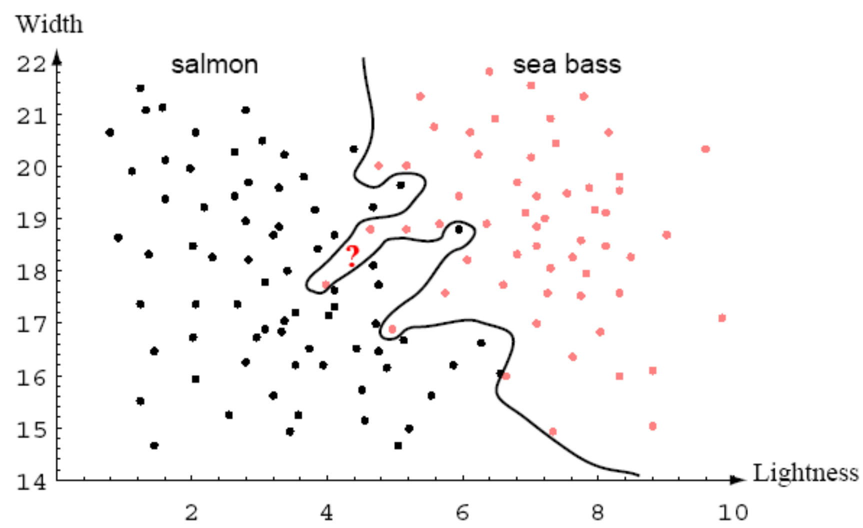


Not good generalization…

- Overly complex models for the fish will lead to decision boundaries that are complicated
- While such a decision <u>may lead</u> to *perfect classification* of our *training samples*, it **<u>would lead to poor performance on future patterns</u>**

Discussion

- What do you think a better decision boundary would be than the one below?



FINAL DECISION

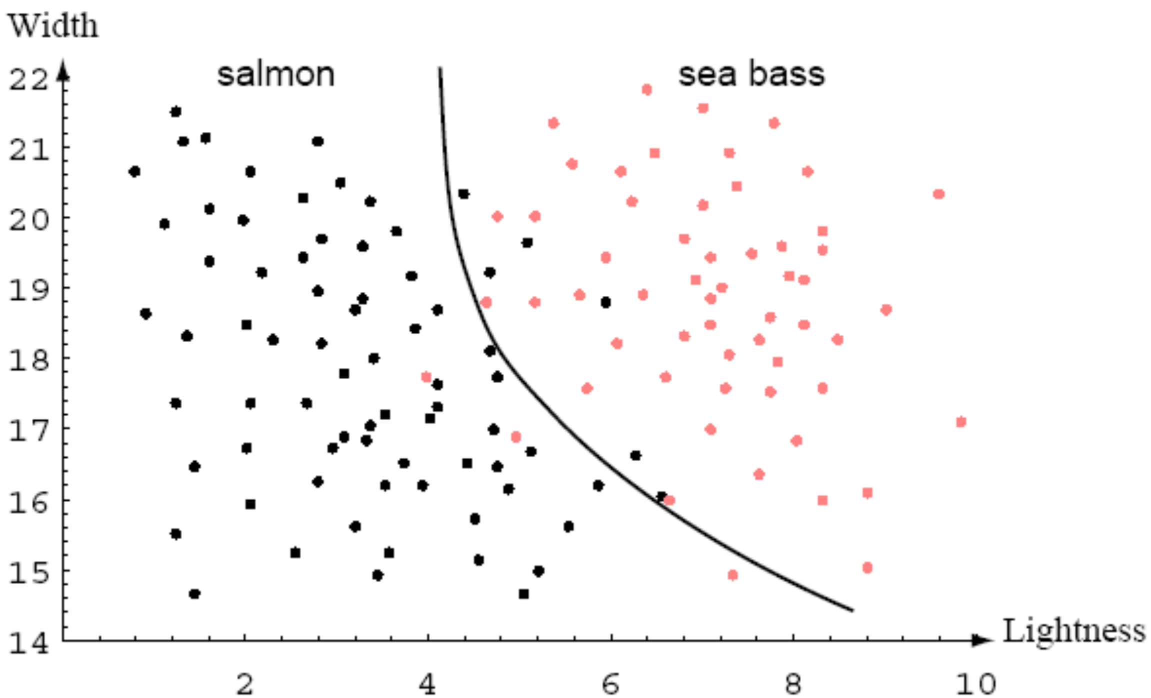


Figure: The decision boundary shown **might represent** the <u>optimal tradeoff</u> **between** *performance on the training set* **and** *simplicity of classifier*.

What can we do?

- Get more training samples => better estimate of the true underlying characteristics (e.g. probability distributions of the categories/classes)
- However, in most PR problems the amount of such data we can obtain easily is <u>often quite limited</u>
- With a vast amount of training data (e.g. continuous feature space)
  - A horrendously complicated decision boundary
  - It would be unlikely to do well on novel patterns!

Central Problems in *Statistical PR*

- Our classifier can have better performance on novel patterns when we may have slightly poorer performance on the training samples

Questions:

- If a *very complex recognizer is unlikely to give* good generalization, precisely how should we quantify and favor simpler classifiers?
- **How would our system automatically determine** that a simple decision boundary (e.g. curve) is preferable to an even simpler (straight line) or a more complicated (complex) boundary?
- If this tradeoff is optimized, can we then *predict how well our system will generalize to new patterns?*

Use of Different Cost Functions

ACTION 1: Separating the fish based on their sex (all females from all species together)

ACTION 2: Cull the damaged fish

=> Different decision tasks

=> Different features may need to be extracted, i.e. more useful for specific actions (*categorization problems*)

Important Notes

- **Classification (at base): the task of recovering the model that generated a set of patterns**

  → *Different classification techniques* **are useful** depending on the type of candidate models themselves (texture vs. appearance-based)

- In **Statistical PR** we focus on the statistical properties of the patterns (generally expressed in probability densities)

> Here the **model for a pattern** *may be* a ***single specific set of features***, though the ***actual pattern sensed*** has been corrupted by some form of random noise

Synthetic Pattern Recognition

- May have a model consisting of some set of crisp logical rules → we employ the methods of SPR*, where **rules** or **grammars** describe our decision*

  > e.g. **Classify** an <u>English sentence</u> as grammatical or not, and here **statistical descriptions** (word frequencies, word correlations, etc.) are **inappropriate**.

Important Notes on Designing Classifiers

- We might wish to favor a small number of features, which might **lead to:**
    1. Simpler **decision regions**, and
    2. A **classifier** easier to train

- We might also wish to have features that are robust, i.e., relatively insensitive to noise or other errors

- In **practical applications** we may need the classifier to:
    1) Act quickly, or
    2) Use **few** electronic components, memory or processing steps

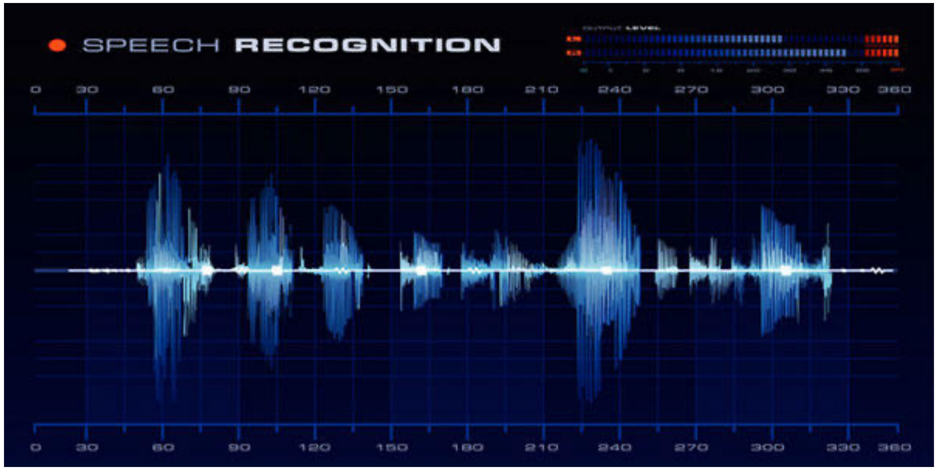
Analysis by Synthesis

- Insufficient training data → need to incorporate knowledge of the problem domain
- The less the training data the more important is such knowledge
  - e.g. how the patterns themselves were produced!

- Analysis by Synthesis
  - Where in the **ideal case** one has a **model** of <u>how each pattern is generated</u>.

Analysis by Synthesis

Examples 1: Speech Recognition



•A "physiological" model (or so-called "motor" model) for production of the utterances can be appropriate

•If this underlying model of production can be determined from the sound, then we can <u>classify the utterance</u> **by how it was produced**

Examples 2:
Image-based Chair Classification!

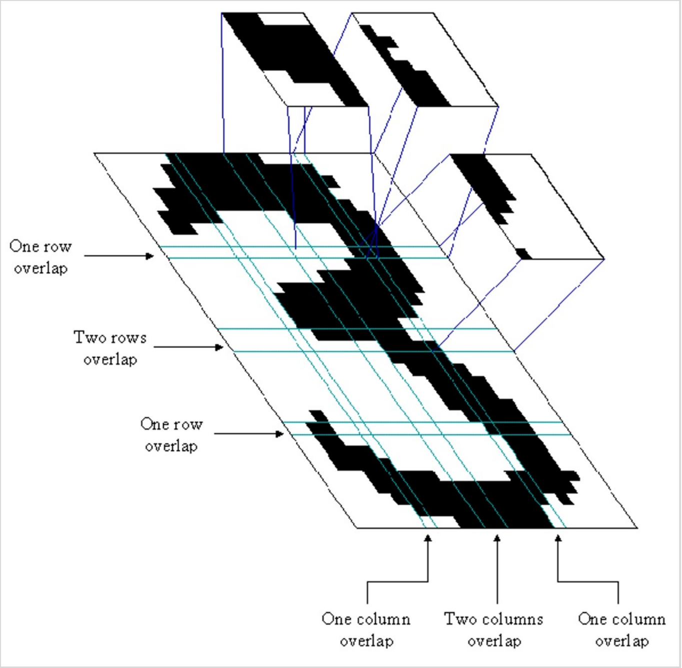
- Plethora in the number of legs, material, shape …
- Is there any sufficient representation (model) that reveals the unity within the class of chairs?
- Perhaps the only such unifying aspect of chairs is ***functional***: a chair is a stable artifact that supports a human sitter, including back support.

Thus, we might try to deduce such functional properties from the image

Real-World Problems
Examples 3: Optical Character Recognition

OCR: is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text.

- In many real world PR systems we need to have *some knowledge about the method of production of the patterns* or their functional use in order to insure a good representation



<u>The goal of the representation is classification, not reproduction</u>

OCR: we can assume that handwritten characters are written as a sequence of strokes →

- Recover a stroke representation from the sensed image
- Deduce the character from the identified strokes

Discussion

- Can you suggest other examples?

Questions?

THANK YOU!