Introduction to Pattern Recognition

Arun Ross

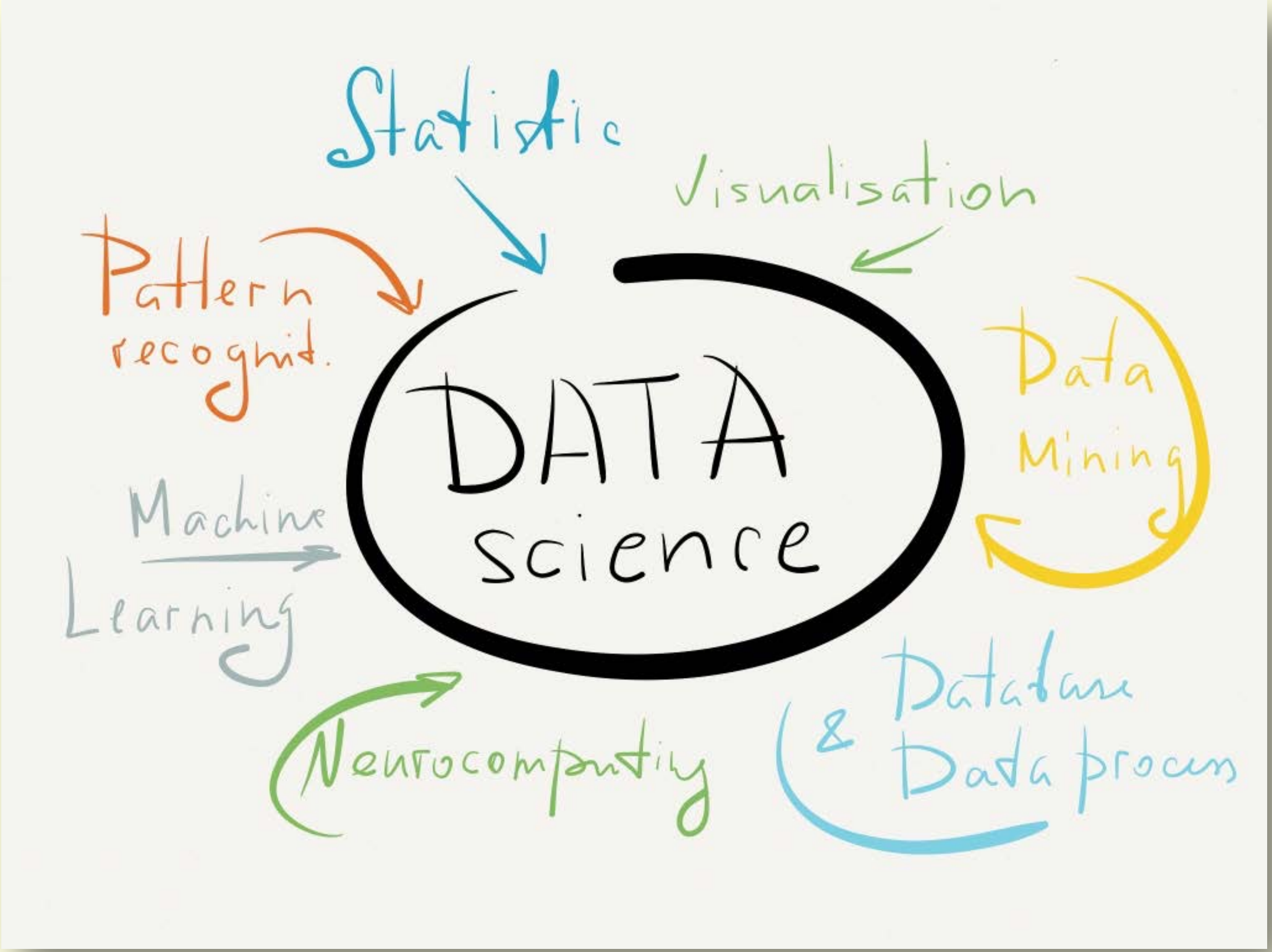
Professor

Michigan State University

rossarun@cse.msu.edu

http://www.cse.msu.edu/~rossarun

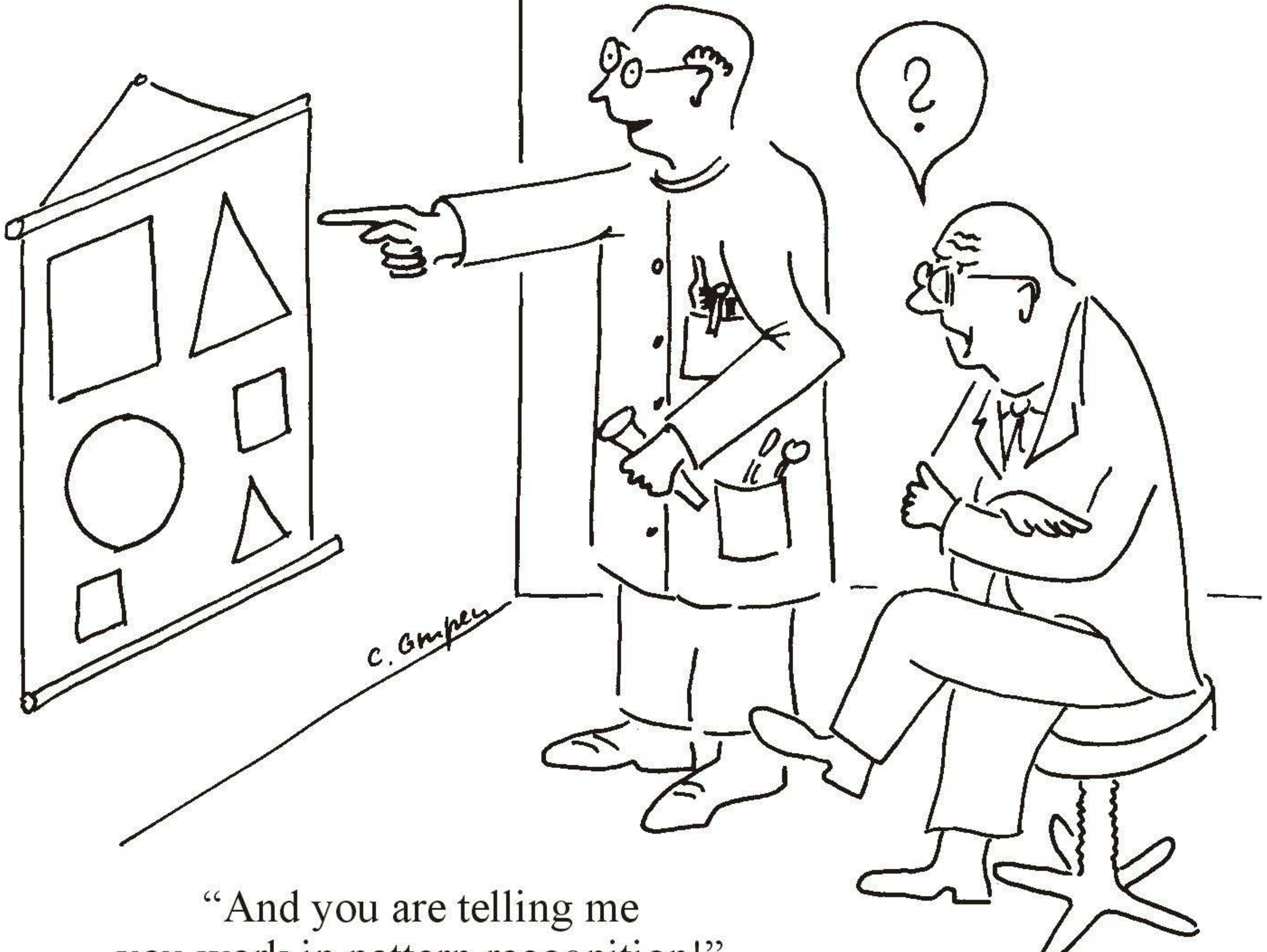
Data Science




DELL
CLARKE
LAMARR
HAMILTON
JOHNSON
HOPPER
WU
LOVELACE
MOLNAR
LAGUNA BEACH, CALIFORNIA - U.S.A.

Patterns!

- "The real power of human thinking is based on recognizing patterns. The better computers get at pattern recognition, the more humanlike they will become." *Ray Kurzweil, NY Times, Nov 24, 2003*
- "The problem of searching for patterns in data is a fundamental one and has a long and successful history." *Christopher Bishop*



"And you are telling me
you work in pattern recognition!"

Pattern Recognition

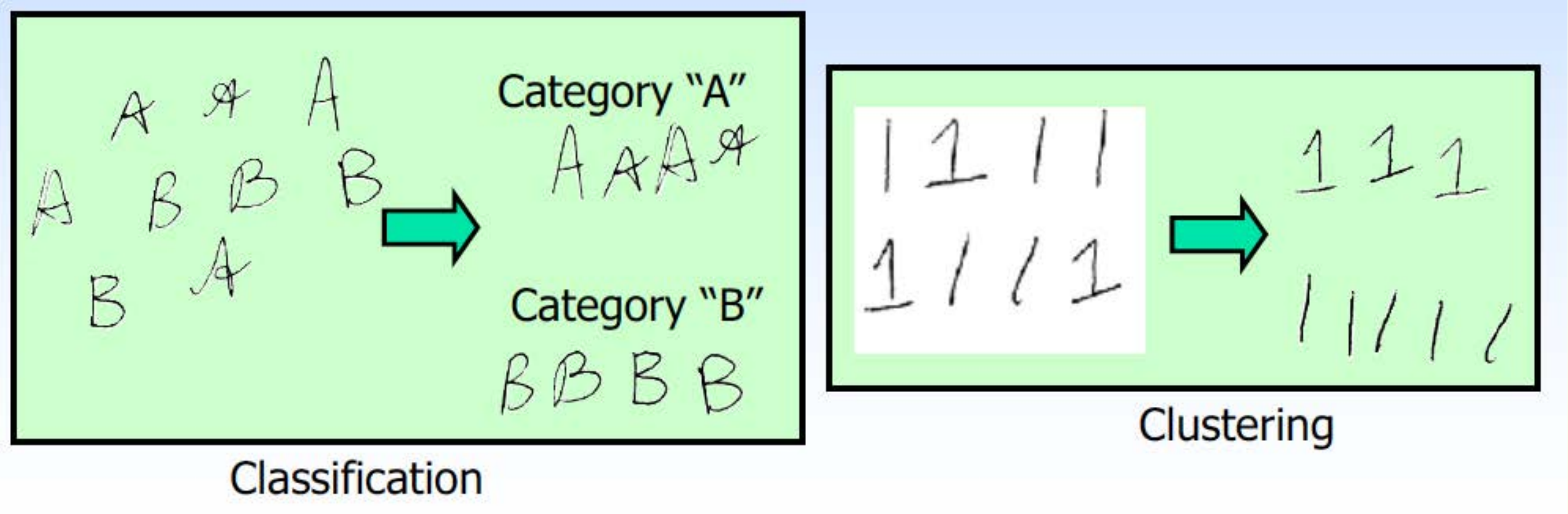
- The act of taking as input sensed data (measurements) and taking an action based on the "category" or "class" of the pattern

What is a Pattern?

- "A pattern is the opposite of a chaos; it is an entity vaguely defined, that could be given a name." (Watanabe)

What is Recognition?

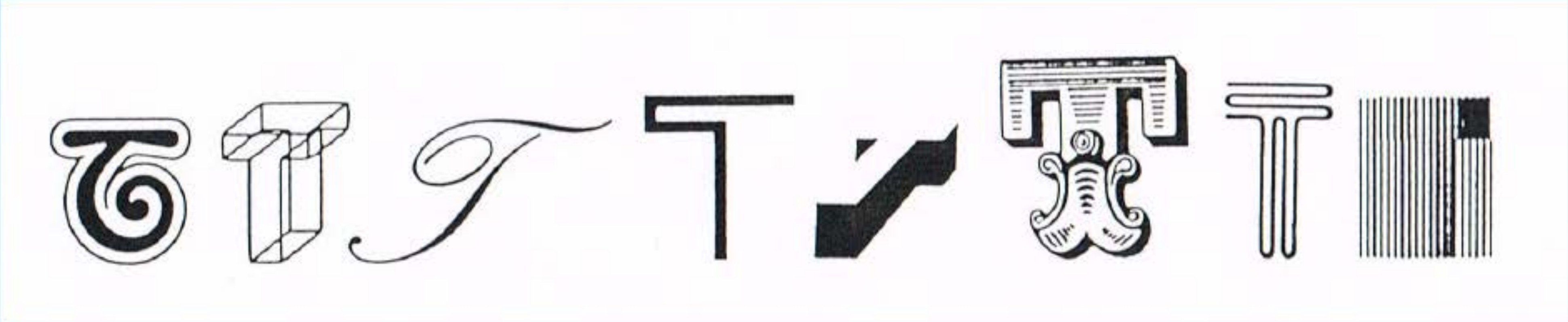
- Identification of a pattern as a member of a category (class) we already know, or we are familiar with
  - **Classification** (known categories)
  - **Clustering** (learning categories)



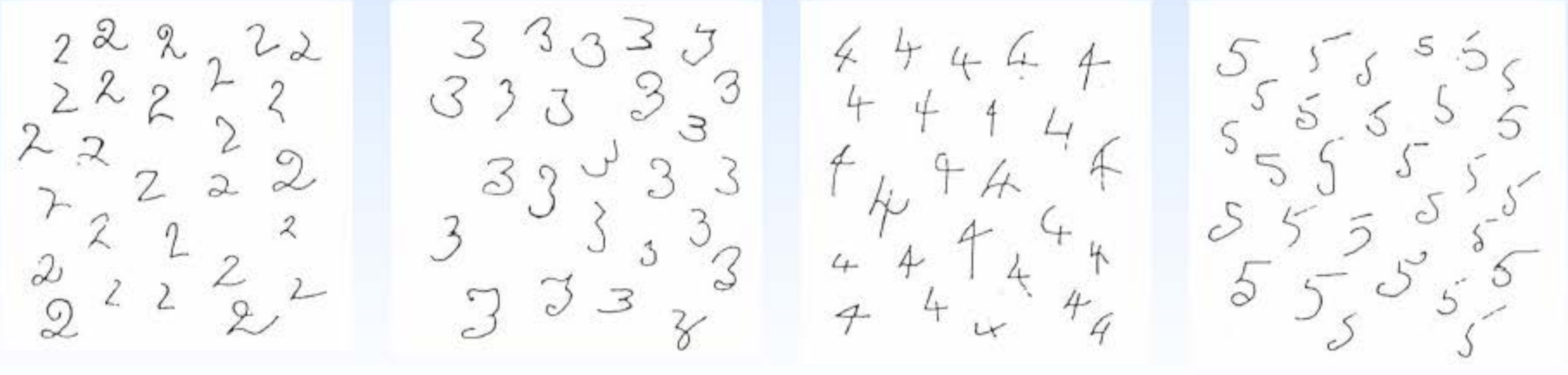
What is a Class?

- A collection of similar (not necessarily identical) objects
- A class is defined by class samples (exemplars, prototypes)
- Intra-class variability
- Inter-class similarity
- How to define similarity?

Intraclass Variability

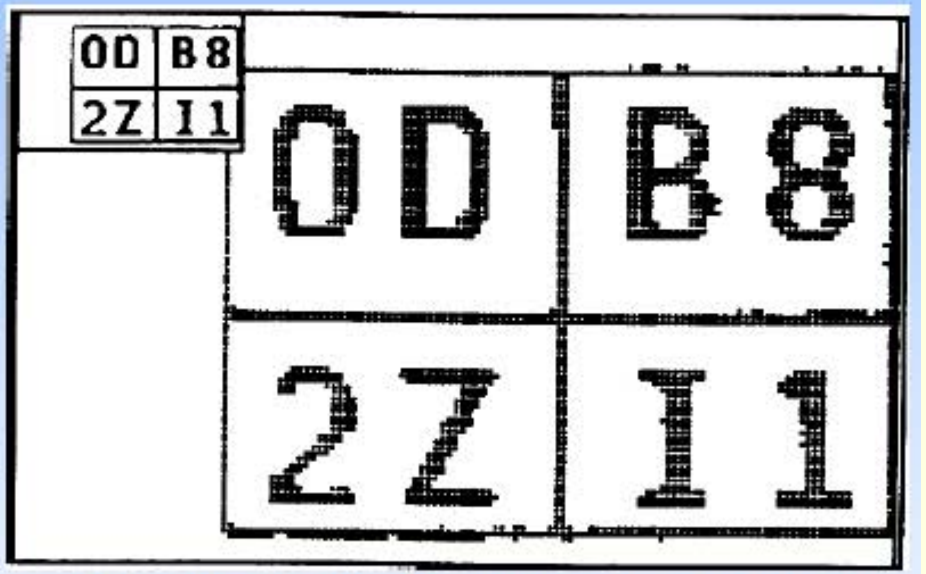


Handwritten numerals



Inter-class Similarity

Identical twins



Characters that look similar

Cat vs Dog (two-class classification)

Supervised Classification

Labeled training samples for classifier design

Unsupervised Classification

Training samples are unlabeled

Pattern Recognition Applications

Problem	Input	Output
Identification and counting of cells	Slides of blood samples, micro-sections of tissues	Type of cells
Industrial inspection (PC boards, IC masks, textiles)	Scanned image (visible, infrared)	Acceptable/unacceptable
Factory automation	3-D images (structured light, laser, stereo)	Identify objects, pose, assembly
Web search	Key words specified by a user	Text relevant to the user
Fingerprint identification	Input image from fingerprint sensors	Owner of the fingerprint, fingerprint classes
Signature recognition (off-line, on-line)	Signature	Financial transactions

Pattern Recognition Applications

Problem	Input	Output
Speech recognition	Speech waveforms	Spoken words, speaker identity
Non-destructive testing	Ultrasound, eddy current, acoustic emission waveforms	Presence/absence of flaw, type of flaw
Medical waveform analysis	EKG, EEG waveforms	Types of cardiac conditions, classes of brain conditions
Remote sensing	Multispectral images	Terrain forms, vegetation cover
Aerial reconnaissance	Visual, infrared, radar images	Tanks, airfields
Character recognition (page readers, zip code, license plate)	scanned image	Alphanumeric characters

Pattern Recognition System

- Challenges
    - Pattern representation
    - Pattern classification
- System design
    - System training or learning
    - System testing or evaluation

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

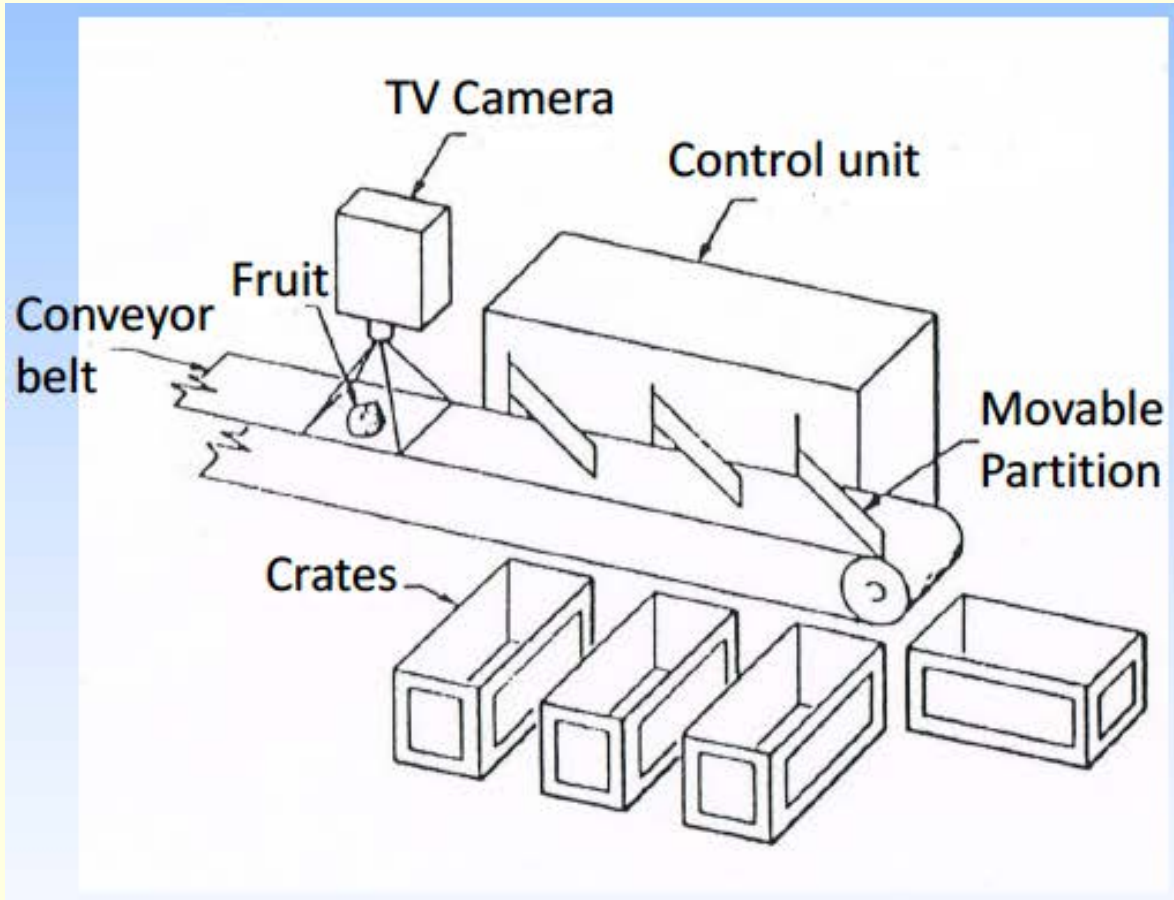
System Performance

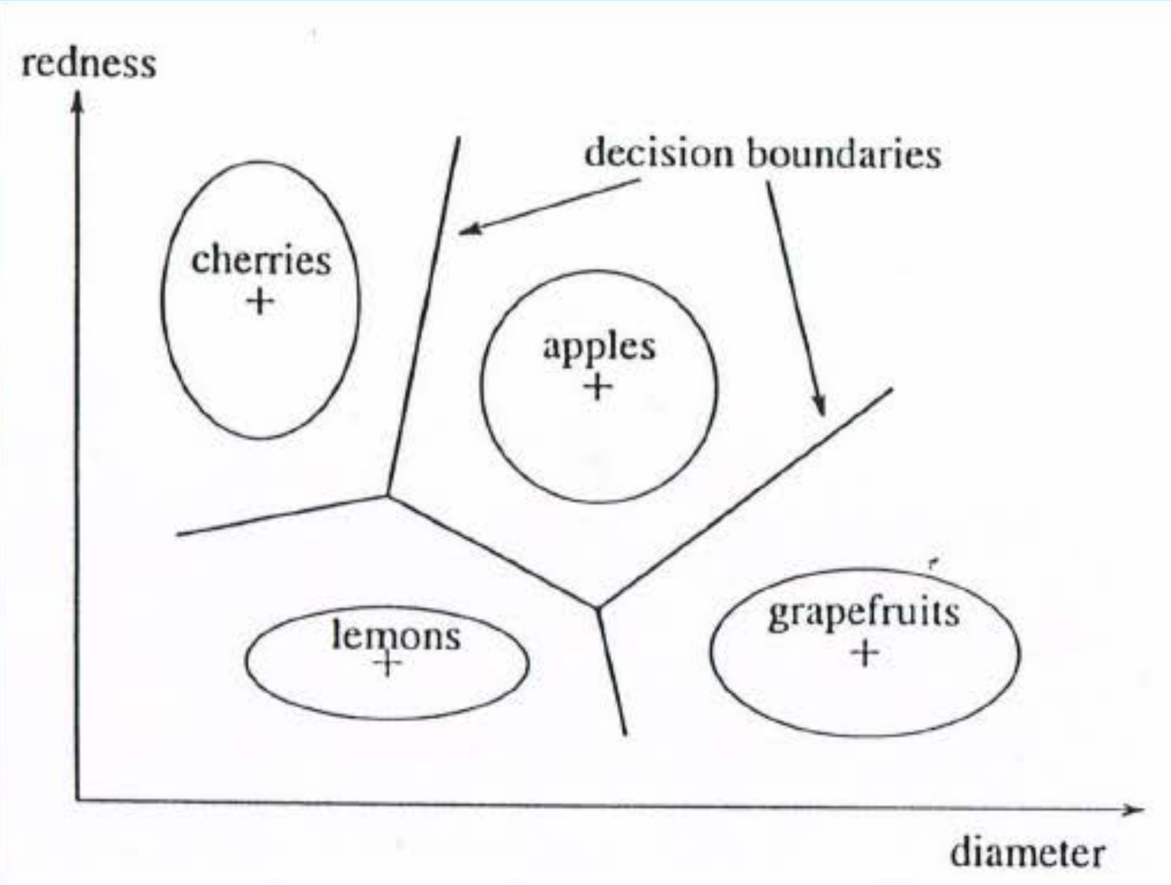
- Error rate; confusion matrix, ROC
- Speed (throughput)
- Cost
- Robustness
- Reject option
- Return on Investment (RoI)

Reject Option

- What if the system encounters a previously unknown class?

Fruit Sorter





Pattern Recognition

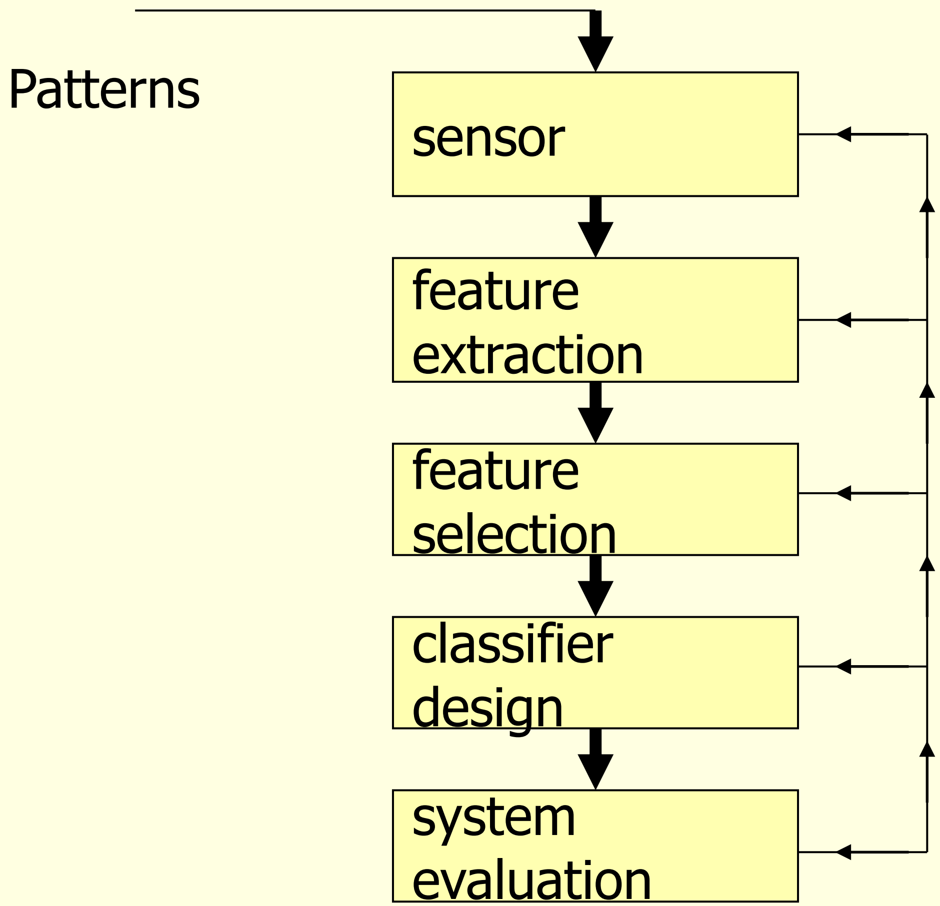
- Assign “patterns” into “classes” using a classifier
- Features: These are measurable quantities obtained from the patterns, and the classification task is based on their respective values.
- Feature vectors: A number of features

$$x_1, \ldots, x_l,$$

constitute the feature vector

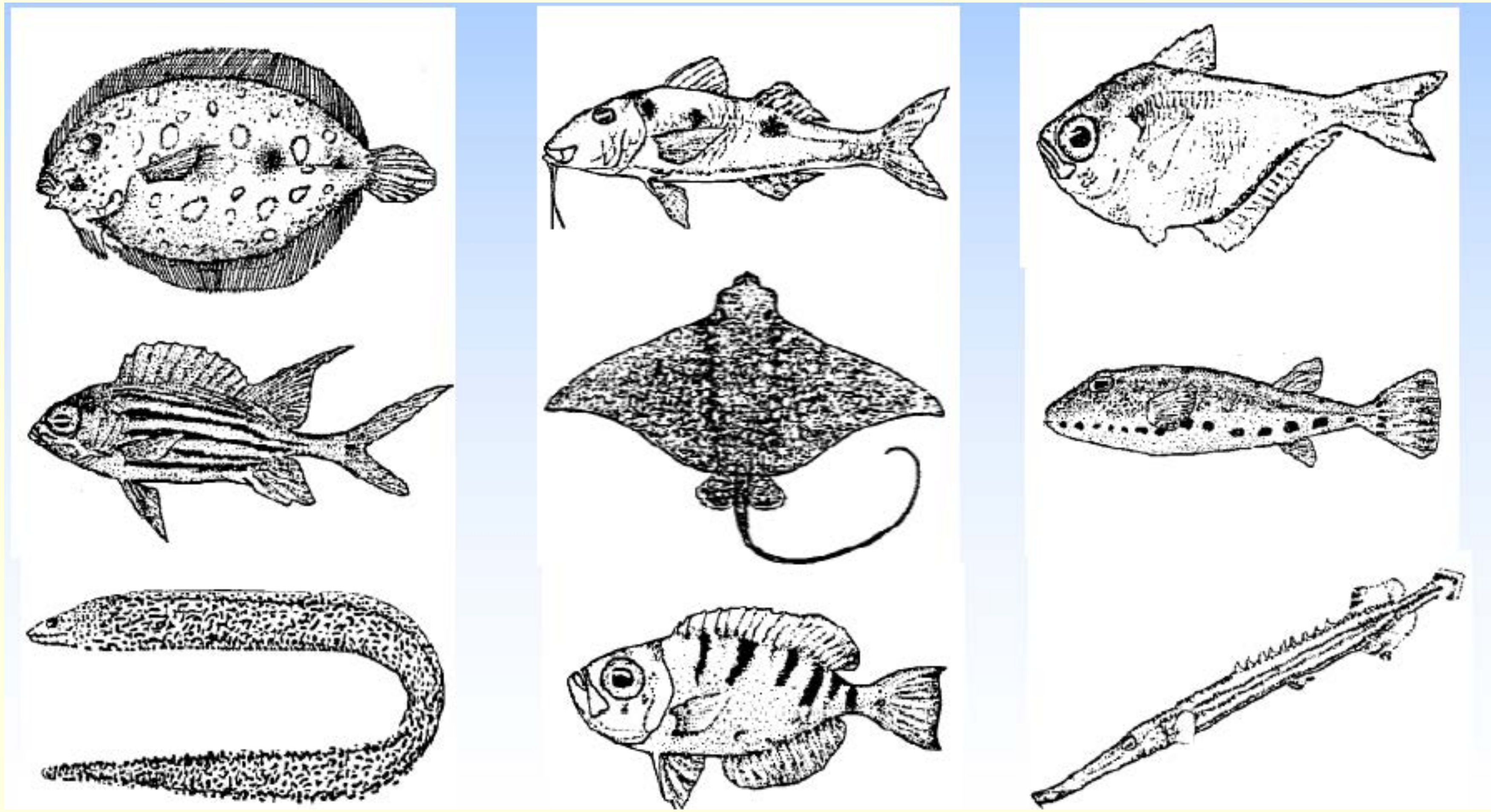
$$\underline{x} = [x_1, \ldots, x_l]^T \in R^l$$

Pattern Recognition System



Fish Sorting

- Cut out each of the fish cards on this page, then follow your teacher's instructions for sorting the fish into categories.
- After you have compared your classification system with your classmates, follow the steps in the fish key below to identify the names of the fish.



http://www-tc.pbs.org/wgbh/nova/education/activities/pdf/2215_reef.pdf

Fish Classification System

Fish key

Step 1
If fish shape is long and skinny…
then go to Step 2

If fish shape is not long and skinny…
then go to step 3

Step 2
If the fish has pointed fins, it is a trumpet fish

If the fish has smooth fins, it is a spotted moray eel

Step 3
If fish has both eyes on top of the head…
then go to step 4

If fish has one eye on each side of the head…
then go to step 5

Step 4
If the fish has long whip-like tail, it is a spotted eagle ray

If the fish has short, blunt tail, it is a peacock flounder

Step 5
If fish has spots…
then go to step 6

If fish does not have spots…
then go to step 7

Step 6
If fish has chin "whiskers," it is a spotted goat fish

If fish does not have chin "whiskers," it is a band-tail puffer

Step 7
If fish has stripes…
then go to step 8

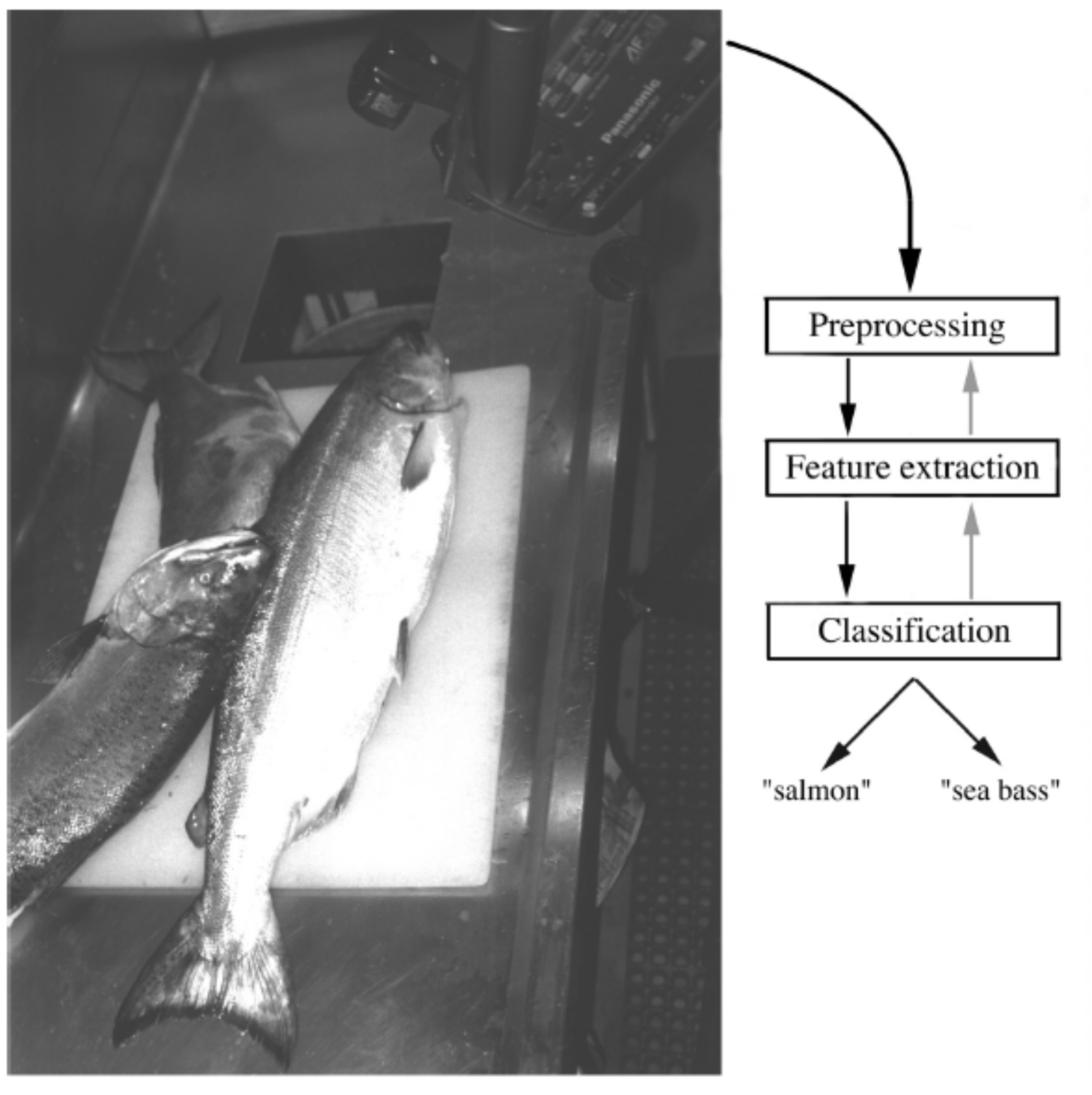
If fish does not have stripes, it is a glassy sweeper

Step 8
If fish has a v-shaped tail, it is a squirrel fish

If fish has a blunt tail, it is a glass-eye snapper

http://www-tc.pbs.org/wgbh/nova/education/activities/pdf/2215_reef.pdf

Fish Classification System



Training Samples: 1-d

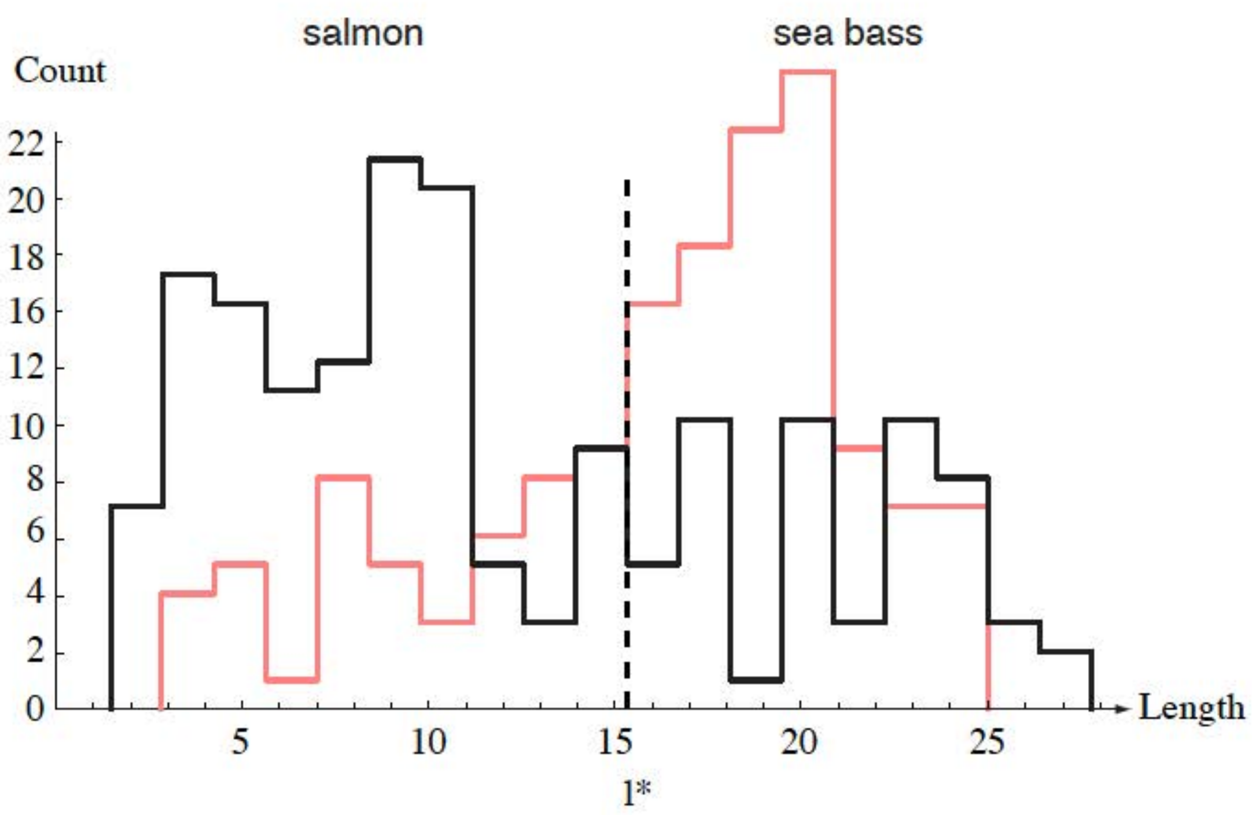


Figure 1.2: Histograms for the length feature for the two categories. No single threshold value l^* (decision boundary) will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value l^* marked will lead to the smallest number of errors, on average.

Training Samples: 1-d

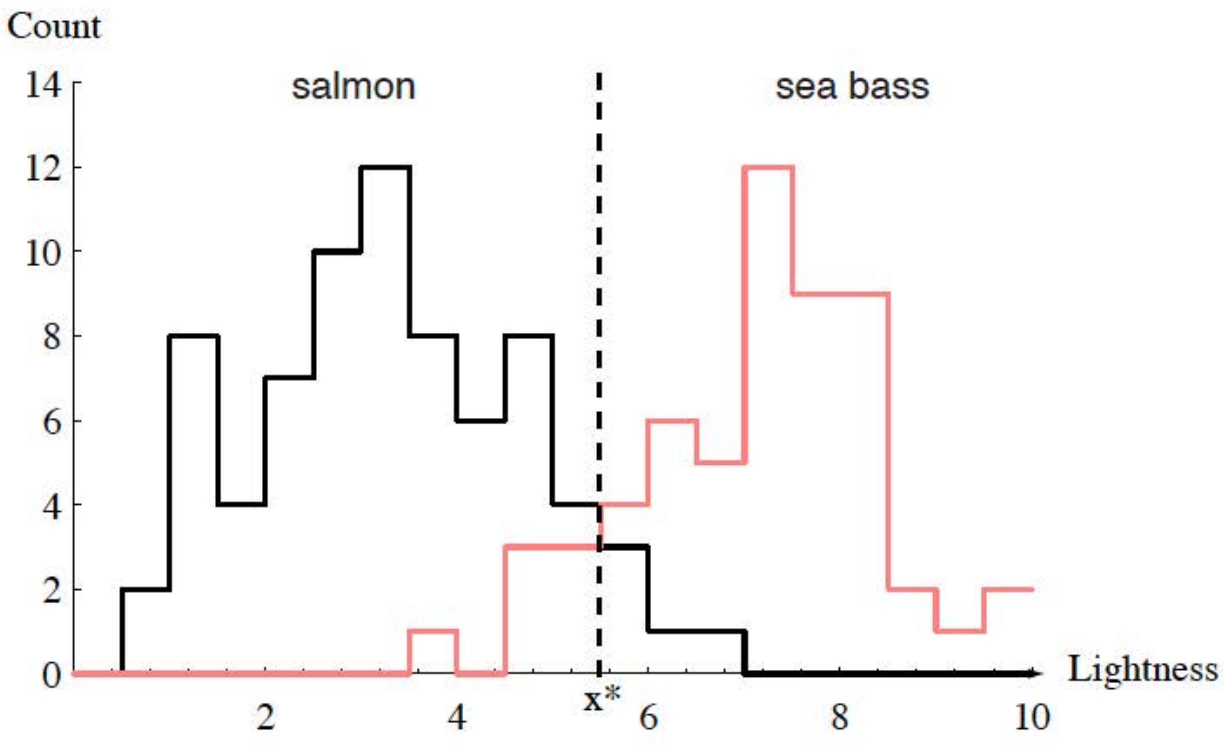


Figure 1.3: Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors, on average.

Training Samples: 2-d

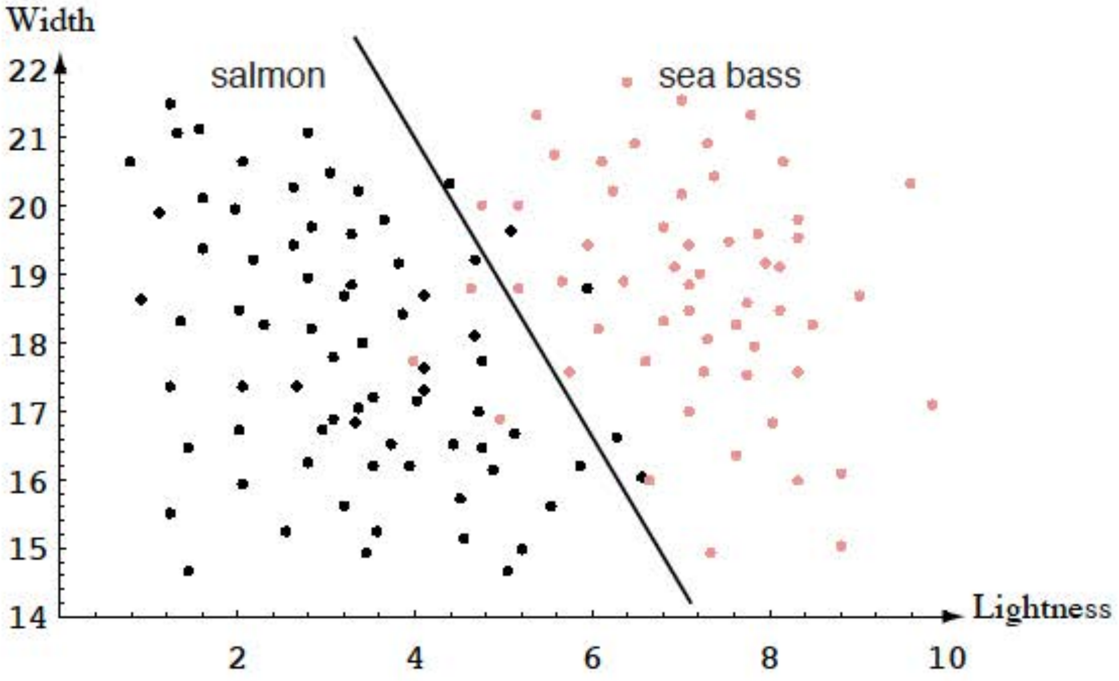


Figure 1.4: The two features of lightness and width for sea bass and salmon. The dark line might serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors.

Complex Decision Boundary

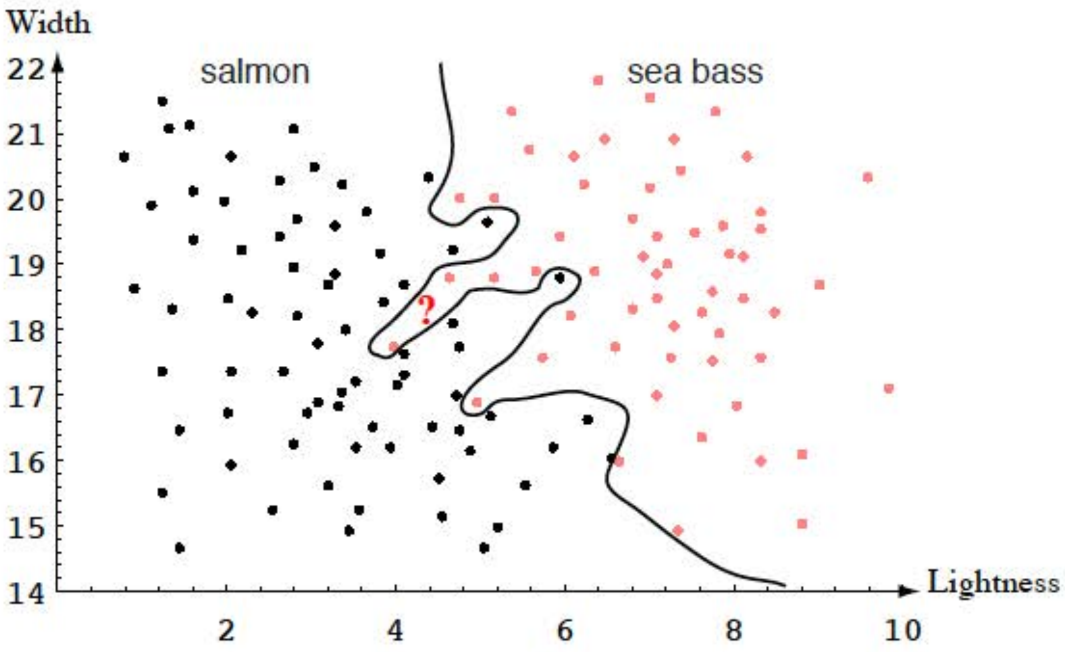


Figure 1.5: Overly complex models for the fish will lead to decision boundaries that are complicated. While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns. The novel test point marked ? is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be misclassified as a sea bass.

Trade-off: Complex vs Simple Decision Boundary

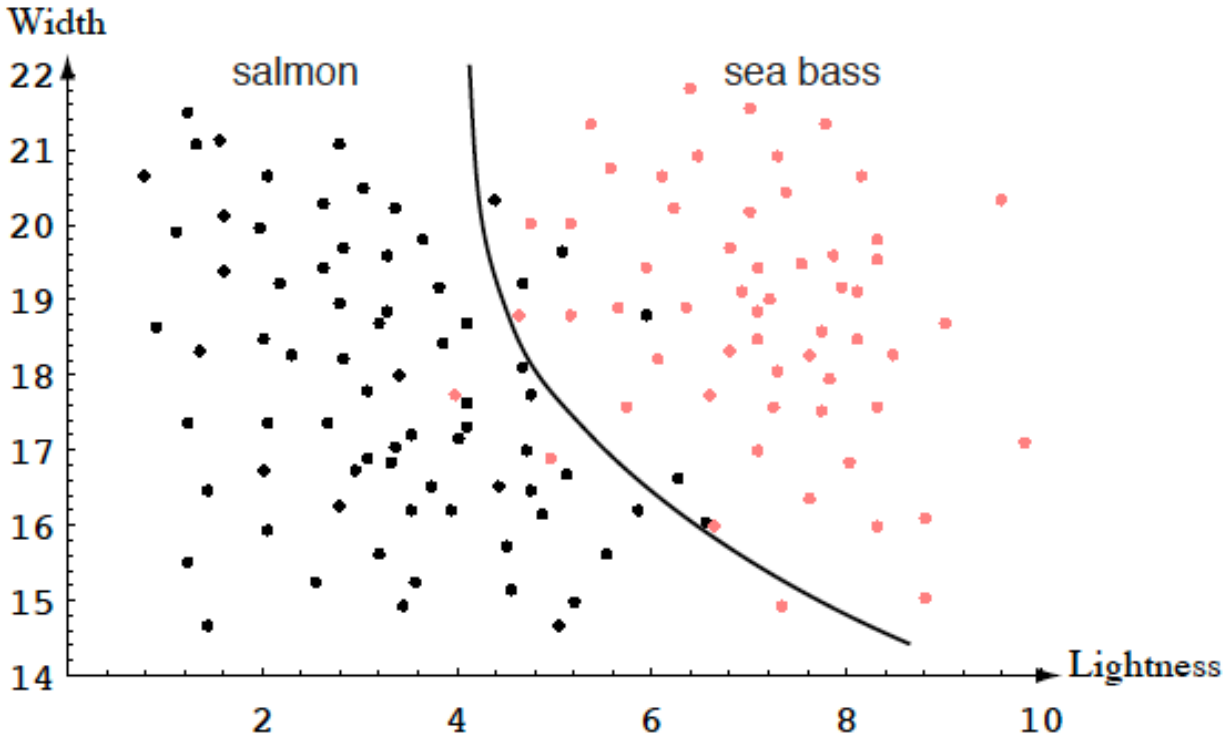


Figure 1.6: The decision boundary shown might represent the optimal tradeoff between performance on the training set and simplicity of classifier.

Occam’s Razor

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

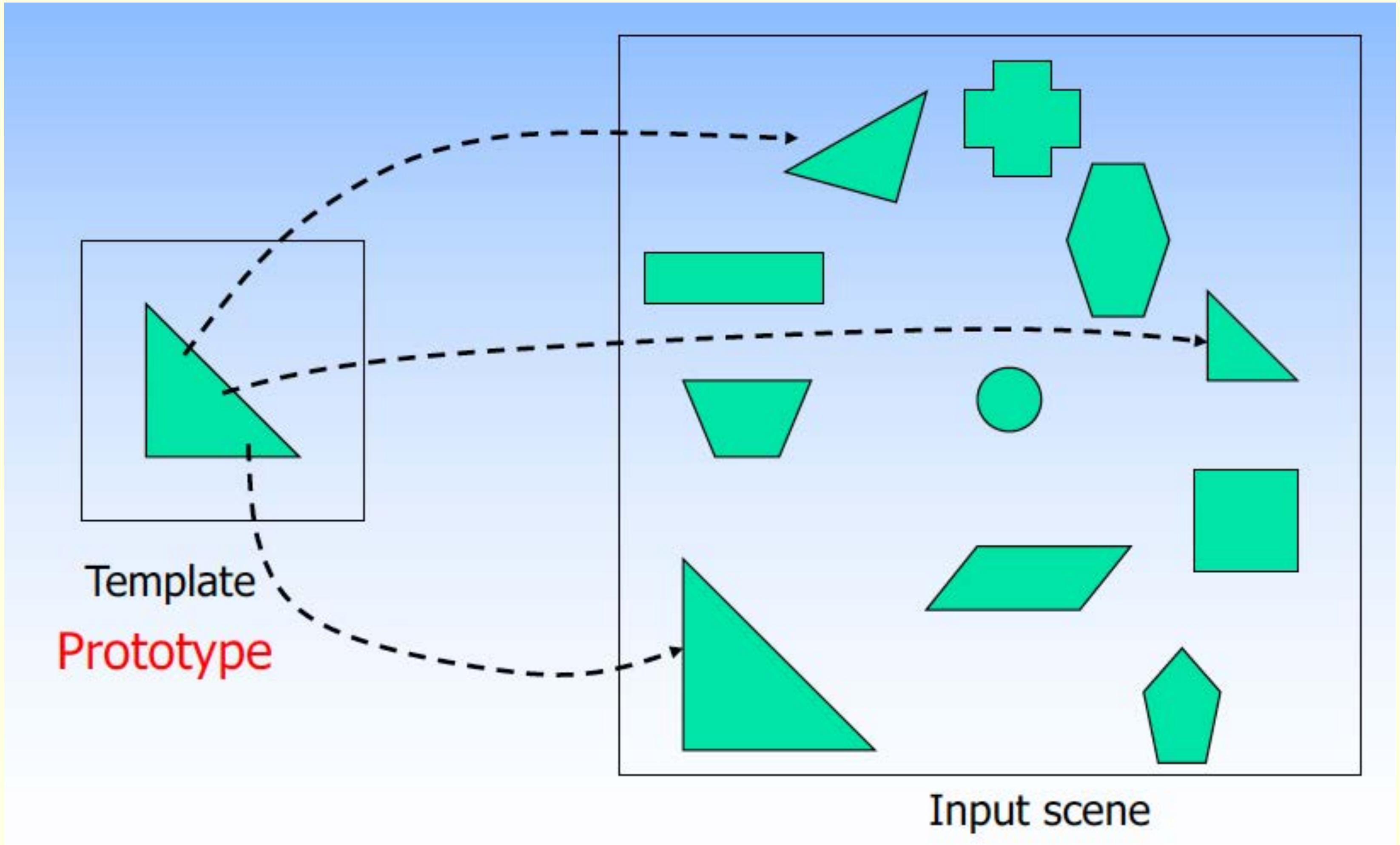
Feature Selection and Extraction

- **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

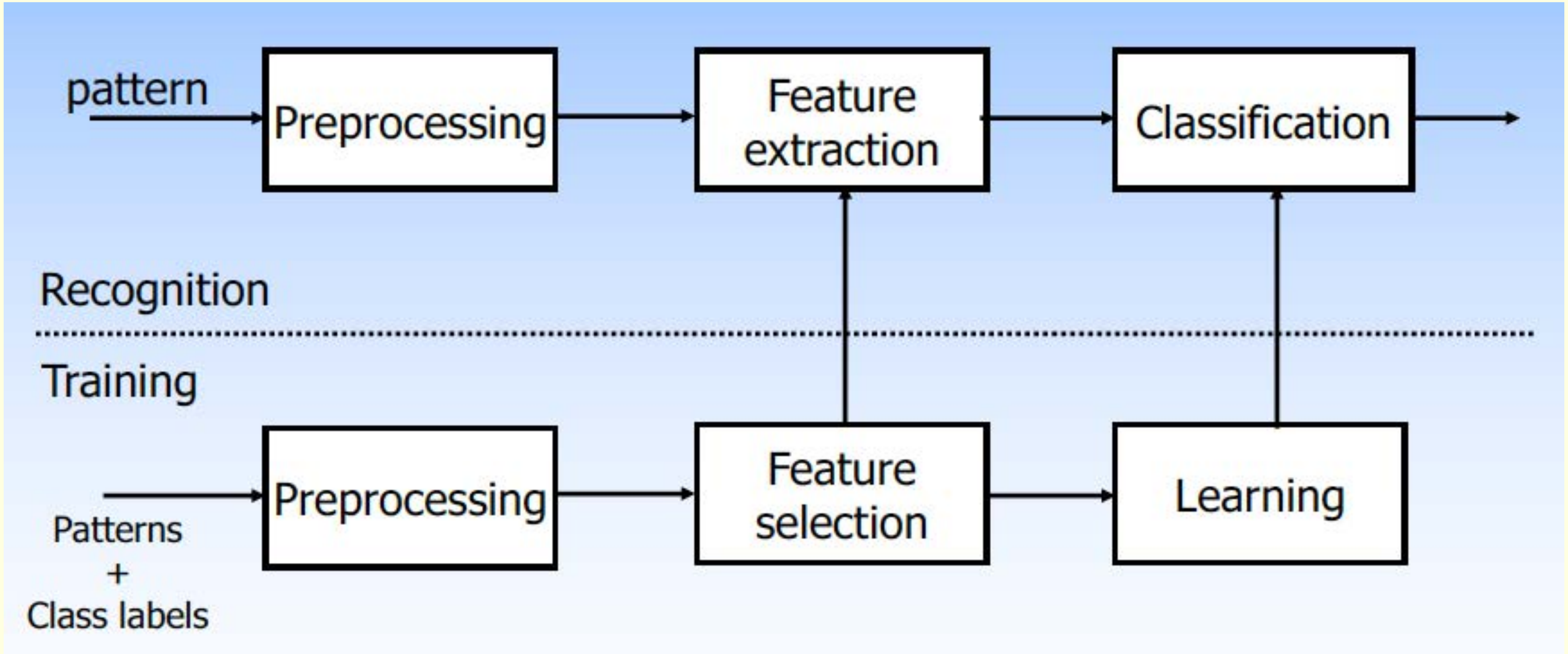
Pattern Recognition Models

- Template matching
  - Class-specific examples
- Statistical (geometric)
  - Class-specific probability density function (pdf)
- Syntactic (structural)
  - Class-specific grammar
- Neural networks

Rigid Template Matching

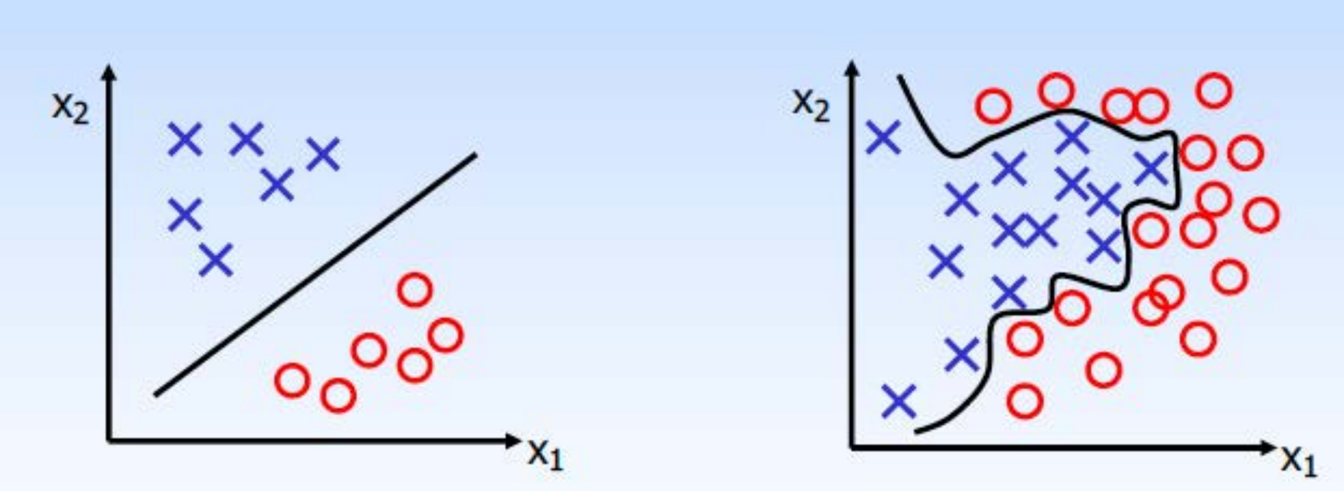


Statistical Pattern Recognition



Representation

- Each pattern is represented as a point in d-dimensional feature space
- Choice of features and their desired invariance properties are domain-specific
- Good representation implies (i) small intra-class variation, (ii) large interclass separation and (iii) simple decision boundary



Probability Density Function

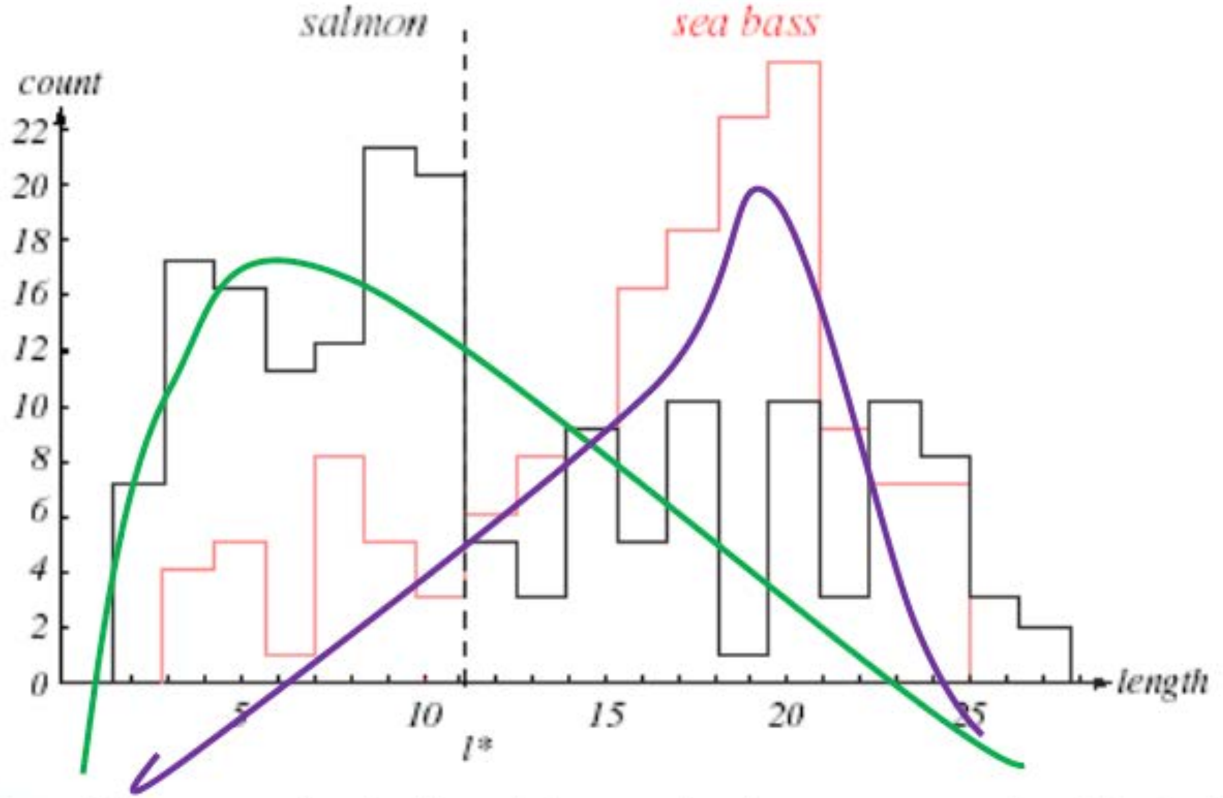


FIGURE 1.2. Histograms for the length feature for the two categories. No single threshold value of the length will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value marked l^* will lead to the smallest number of errors, on average.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

Probability Density Function

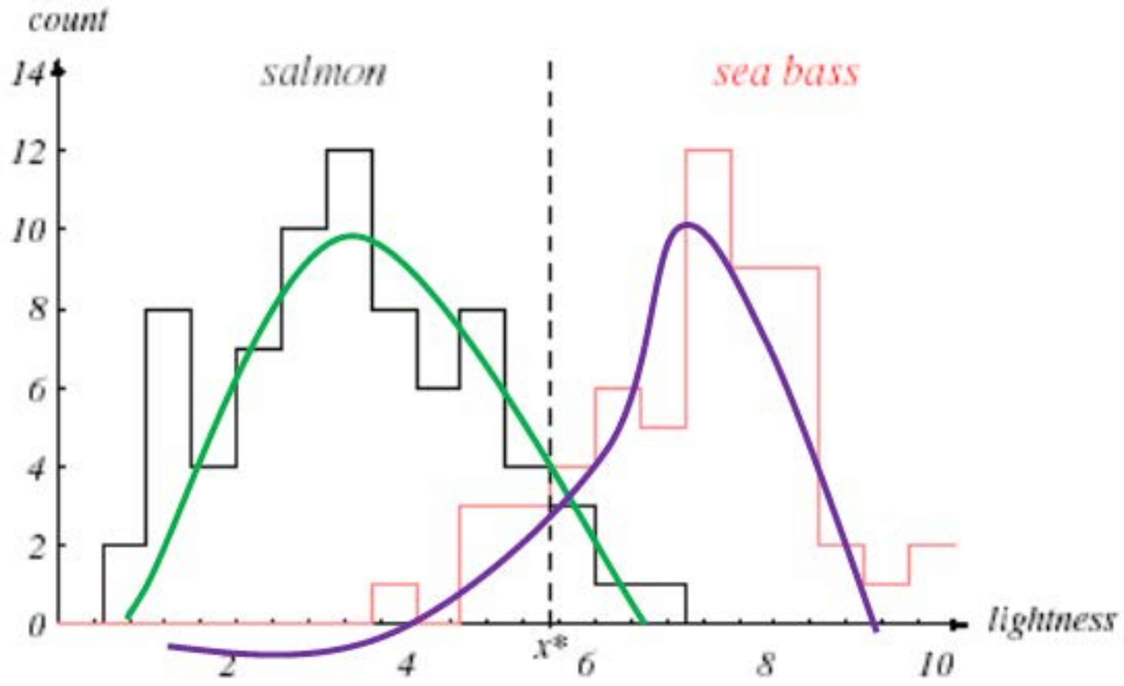


FIGURE 1.3. Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors,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

Probability Density Function

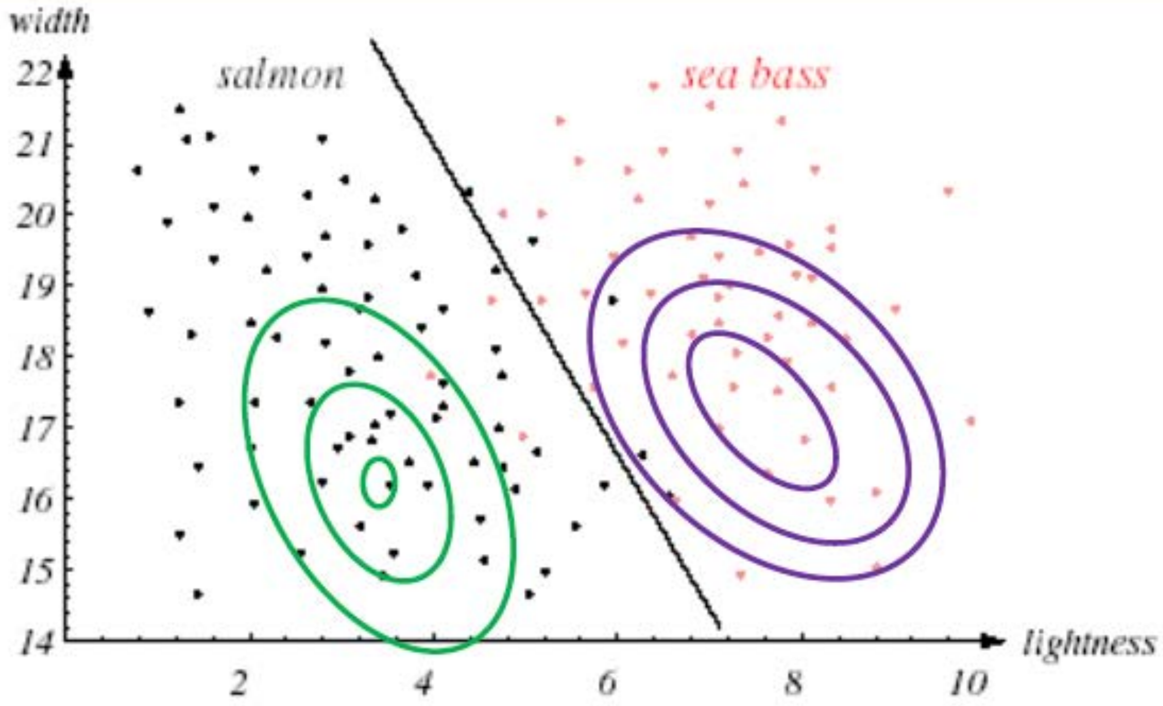
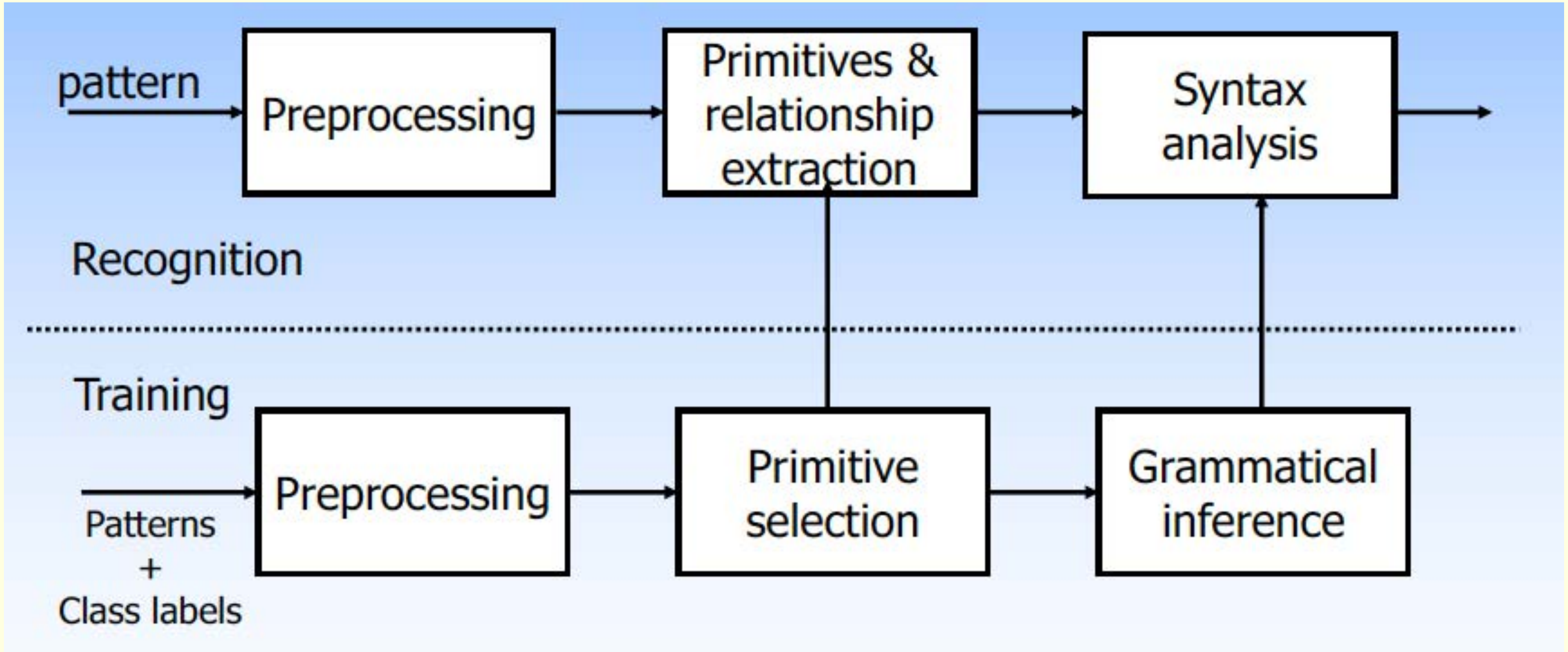


FIGURE 1.4. The two features of lightness and width for sea bass and salmon. The dark line could serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors. From: Richard O.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

Structural Pattern Recognition

- Instead of describing an object in terms of a feature vector, describe it by its structure
- Complex objects are represented in terms of simple primitives (shapes) and their relationship
- Parts-based representation (represent face as eyes, mouth, nose,…)

Syntactic Pattern Recognition



Neural Networks

- Massive parallelism essential for complex recognition tasks (speech & image recognition)
  - Humans take only ~200ms. for most cognitive tasks; this suggests parallel computation in human brain
- Biological networks achieve excellent recognition performance via dense interconnection of simple computational elements (neurons)
  - Number of neurons: $\sim 10^{10} - 10^{12}$
  - Number of interconnections/neuron: $\sim 10^{3} - 10^{4}$
  - Total number of interconnections: $\sim 10^{14}$

Classifier Fusion

- Combine the evidence from component recognizers; also known as classifier combination, mixture of experts, evidence accumulation

"Evolution" of Field

- **Fisher Linear Discriminant (1936)**
- **Perceptron, Rumelhart (1958)**
- **Adaptive multilayer networks, Widrow (1960s)**
- **Backpropagation learning algorithm, Werbos (1974)**
- **Artificial Intelligence (AI), McCarthy (1956)**
- **Pattern recognition**
- **Artificial neural networks**
- **Data mining**
- **Machine learning**
- **Knowledge discovery, expert systems**
- **Deep networks, convolution neural networks**

Recognition, Classification,
Clustering, Regression

Key Concepts

- Pattern class
- Representation, feature set
- Feature selection
- Feature extraction
- Linear transformation (PCA, LDA)
- Feature invariance
- Preprocessing
- Segmentation
- Training set
- Validation set
- Test set
- Error rate
- Reject rate
- Curse of dimensionality

Key Concepts

- Supervised learning
- Decision boundary
- Classifier
- Unsupervised learning
- Clustering
- Density Estimation
- Cost of misclassification/Risk
- Feature space partitioning
- Generalization (overfitting)
- Contextual information
- Combination of classifiers
- Prior knowledge

Classifiers

- Logistic Regression
- Bayes Classifier
- Nearest Neighbor
- Support Vector Machines
- Decision Trees
- Boosted Trees
- Random Forest
- Neural Networks