National University of Sciences and Technology (NUST)

School of Electrical Engineering and Computer Science (SEECS)

Digital Image Processing

Khawar Khurshid, 2012

Intro to Pattern Recognition

What is a Pattern?

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

PR Definitions from Literature

"The assignment of a physical object or event to one of several pre-specified categories" – *Duda and Hart*

Pattern Recognition is concerned with answering the question "*What is this?*" *–Morse*

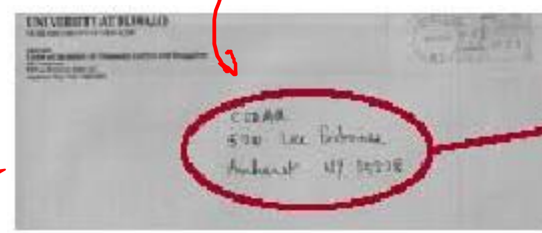
Examples of PR Applications

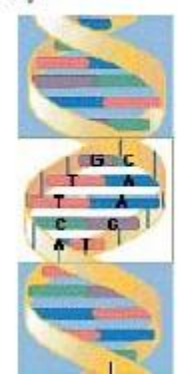
Optical Character Recognition (OCR)

- Sorting letters by postal code.
- Reconstructing text from printed materials.
- Handwritten character recognition

Analysis and identification of human patterns (Biometric Classification)

- Face recognition
- Handwriting recognition
- Finger prints and DNA mapping
- Iris scan identification



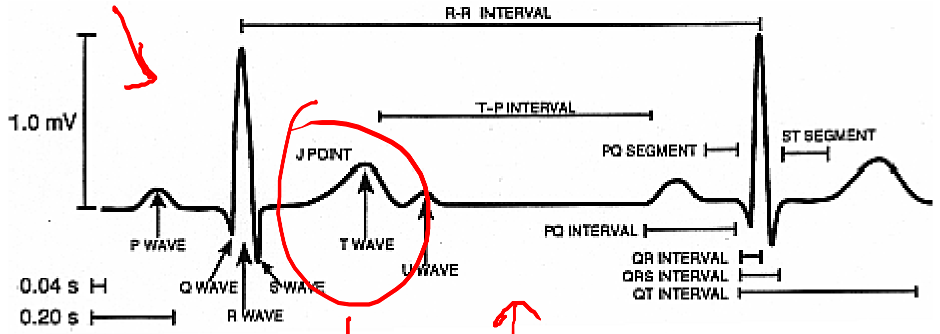


Heart Rate

Examples of Pattern Recognition Problems

Computer aided diagnosis

- Medical imaging, EEG, ECG signal analysis
- Designed to assist (not replace) physicians



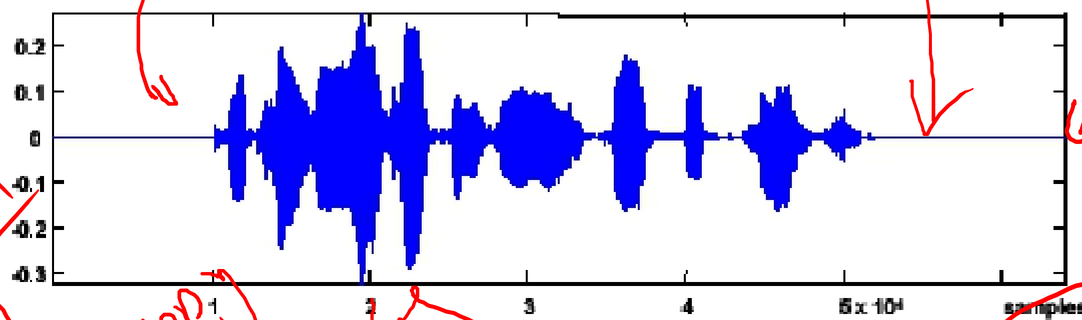
Prediction systems

- Weather forecasting (based on satellite data).
- Analysis of seismic patterns

Earthquake

Other Examples

Speech and voice recognition/speaker identification.



Sentiment analysis

Happy

Google

This is a cat

Banking and insurance applications

– Credit cards applicants classified by income, credit worthiness, mortgage amount, # of dependents, etc.

– Car insurance (pattern including make of car, # of accidents, age, gender, driving habits, location, etc).

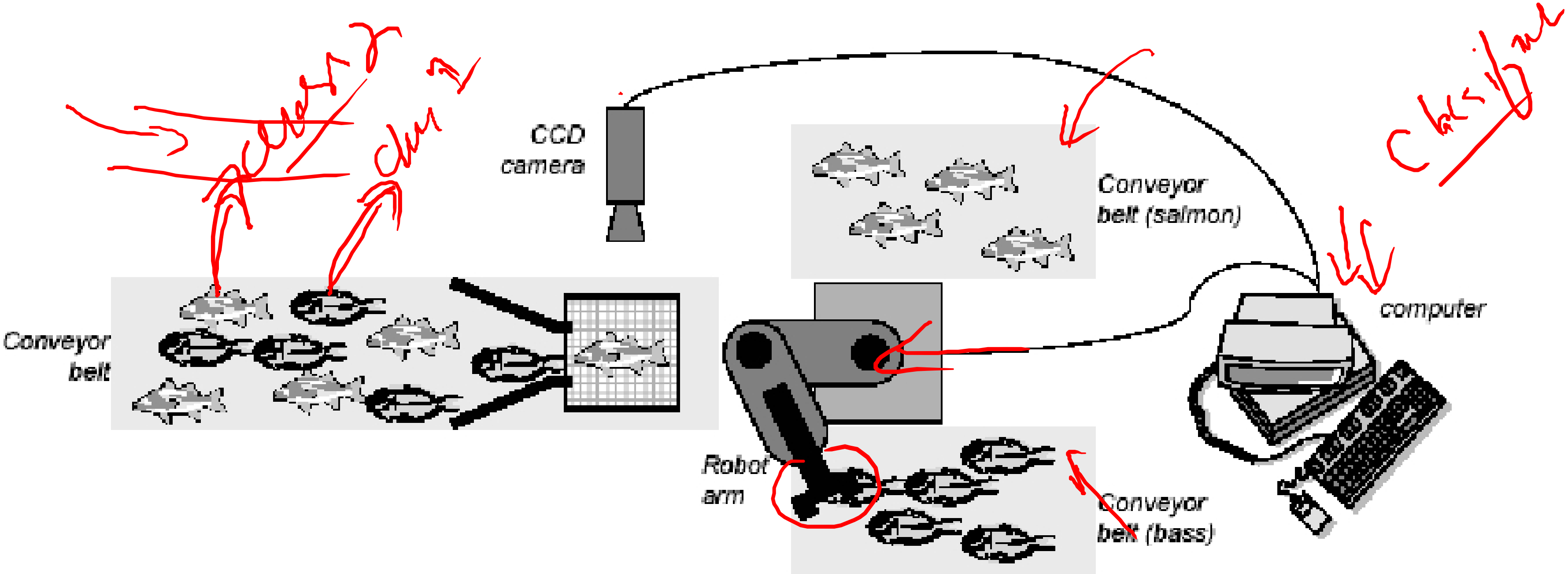
A Pattern Classification Example

A fish processing plant wants to automate the process of sorting incoming fish according to species (salmon or sea bass)

fishing net

The automation system consists of

- A conveyor belt for incoming products
- A vision system with an overhead CCD camera
- A computer to analyze images and control the robot arm



Computer vision, Image Processing and PR

ultra 1K Camera ~~600 fps~~

Real world

Image Acquisition by sensor

Image enhancement and restoration

Filtering

Feature Extraction

Edge detection

Texture

Interest points

3

5

Segmentation

Impose some order on group of pixels to separate them from each other

where?

w

which?

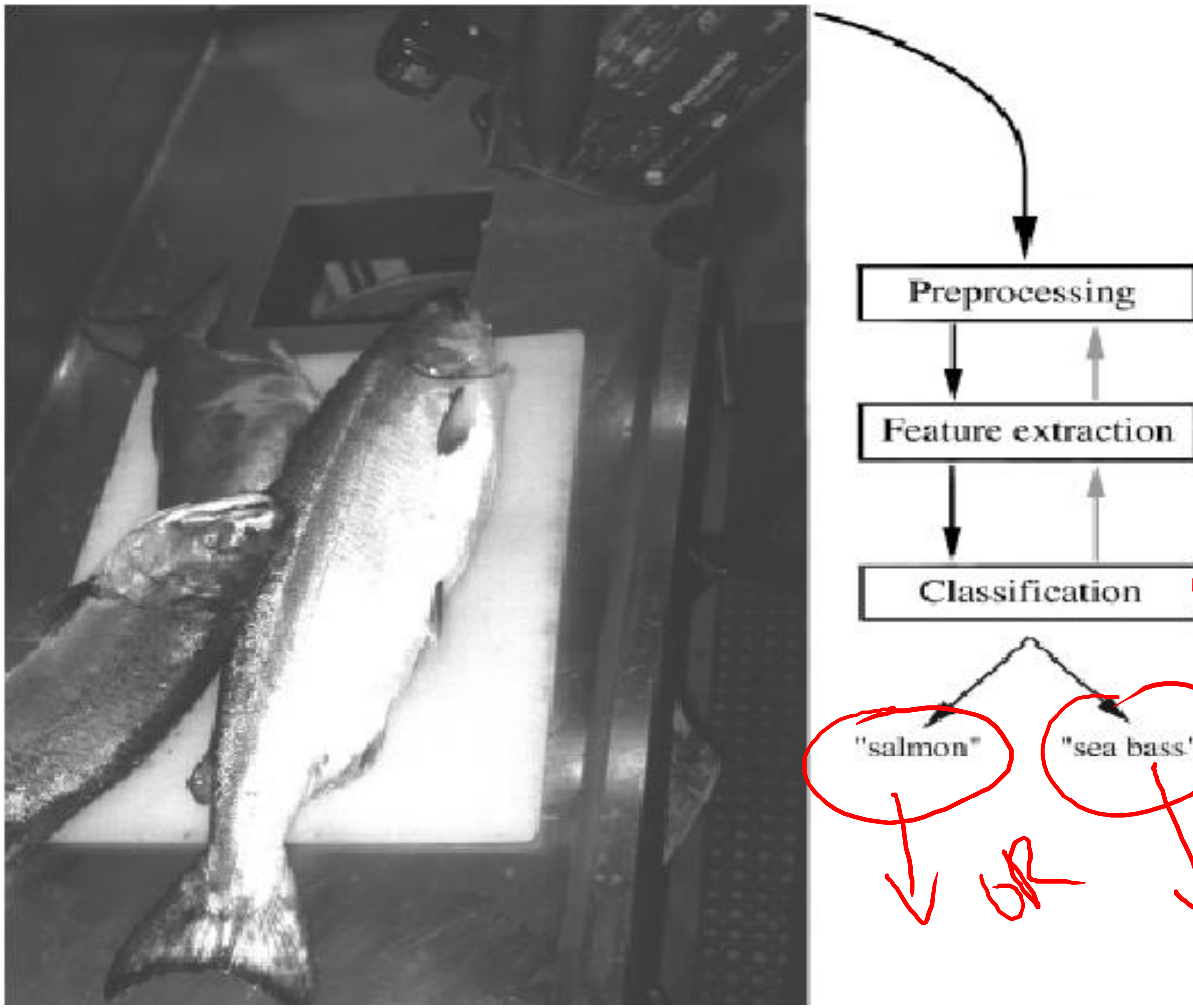
Classification and Recognition

Classification and interpretation of objects based on selected features

Recognize objects using probabilistic techniques

A Pattern Classification Example

"Sorting incoming Fish on a conveyor according to species using optical sensing"



OR

Problem Analysis

Set up a camera and take some training samples to extract features

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

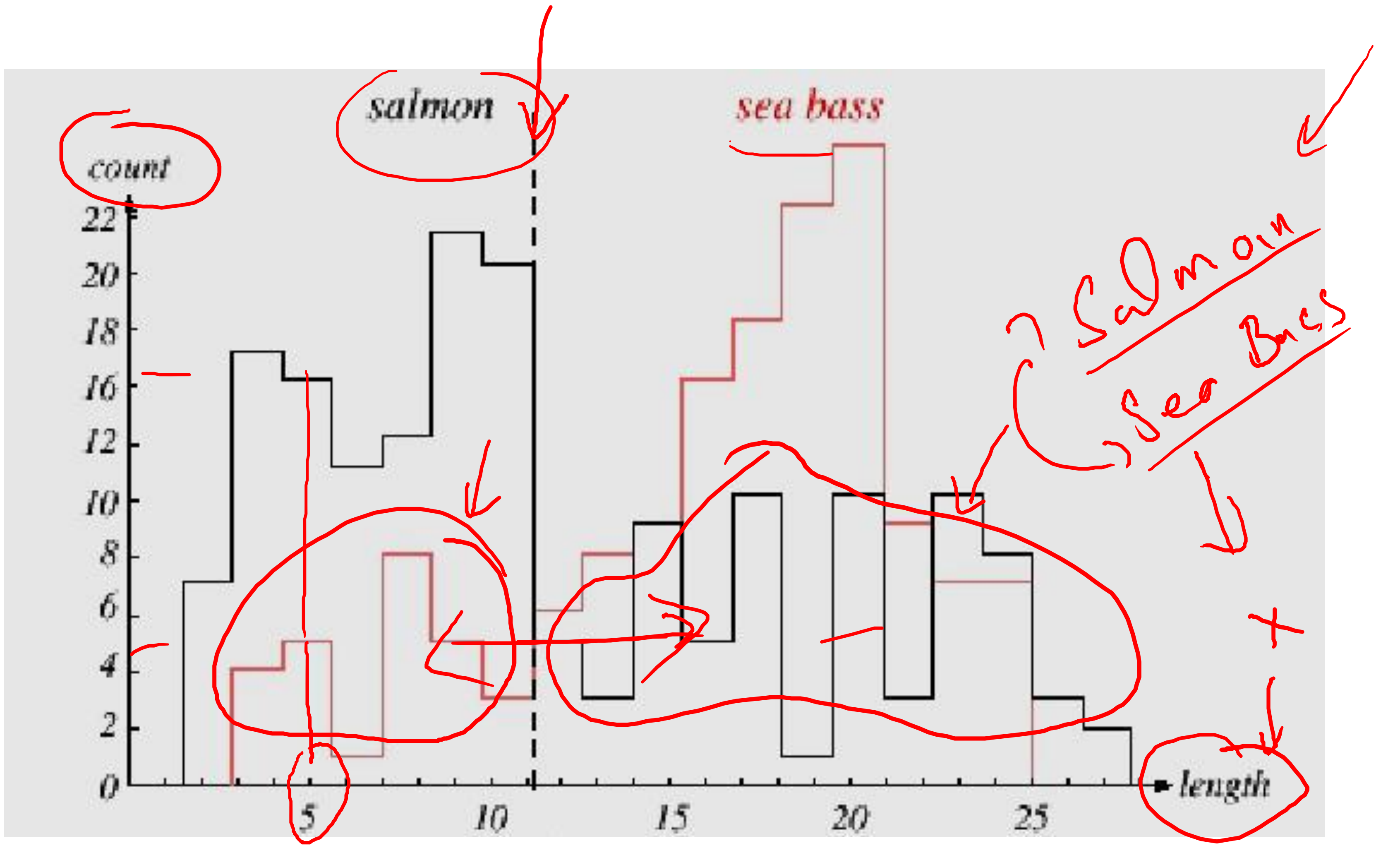
– This is the set of all suggested features to explore for use in our classifier!

Purpose:

To classify the future samples based on the data of extracted features from the training samples

Selection Criterion

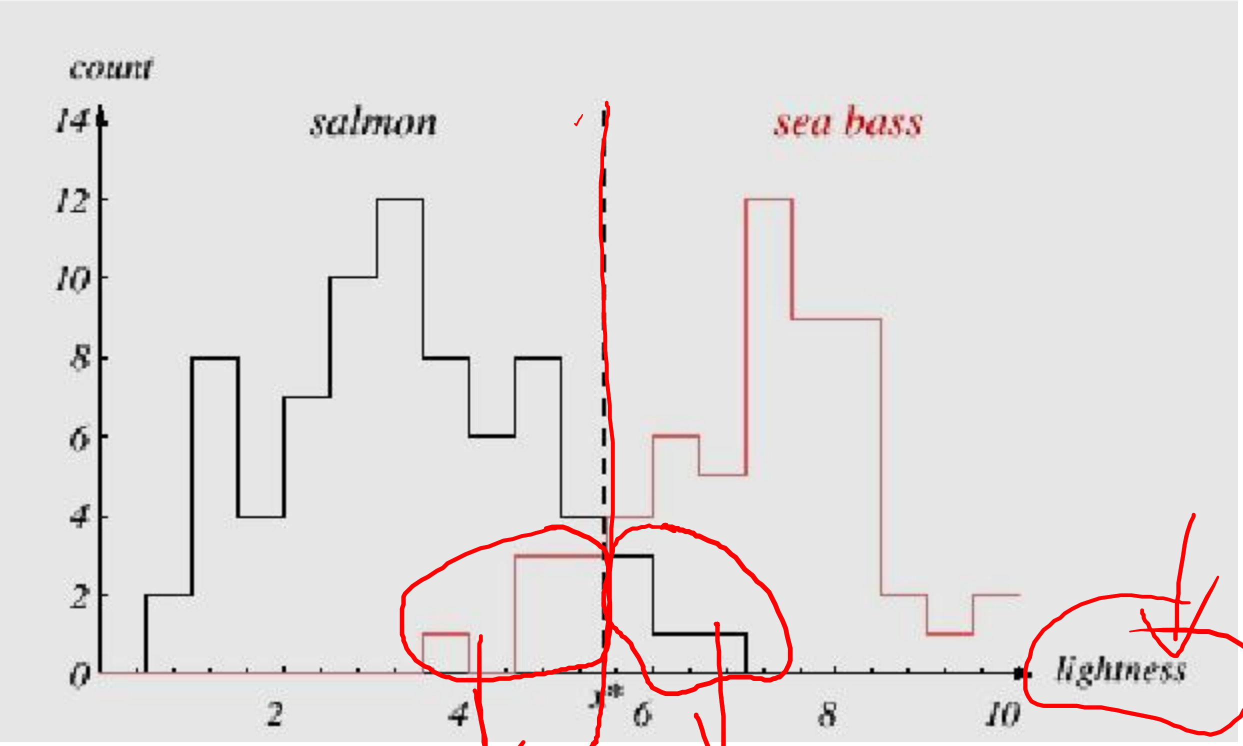
Select the length of the fish as a possible feature for discrimination



Selection Criterion

The length is a poor feature alone!

Select the lightness as a possible feature.



Length or Lightness (weight), which one is a better feature? Lightness

No value of either feature will "classify" all fish correctly

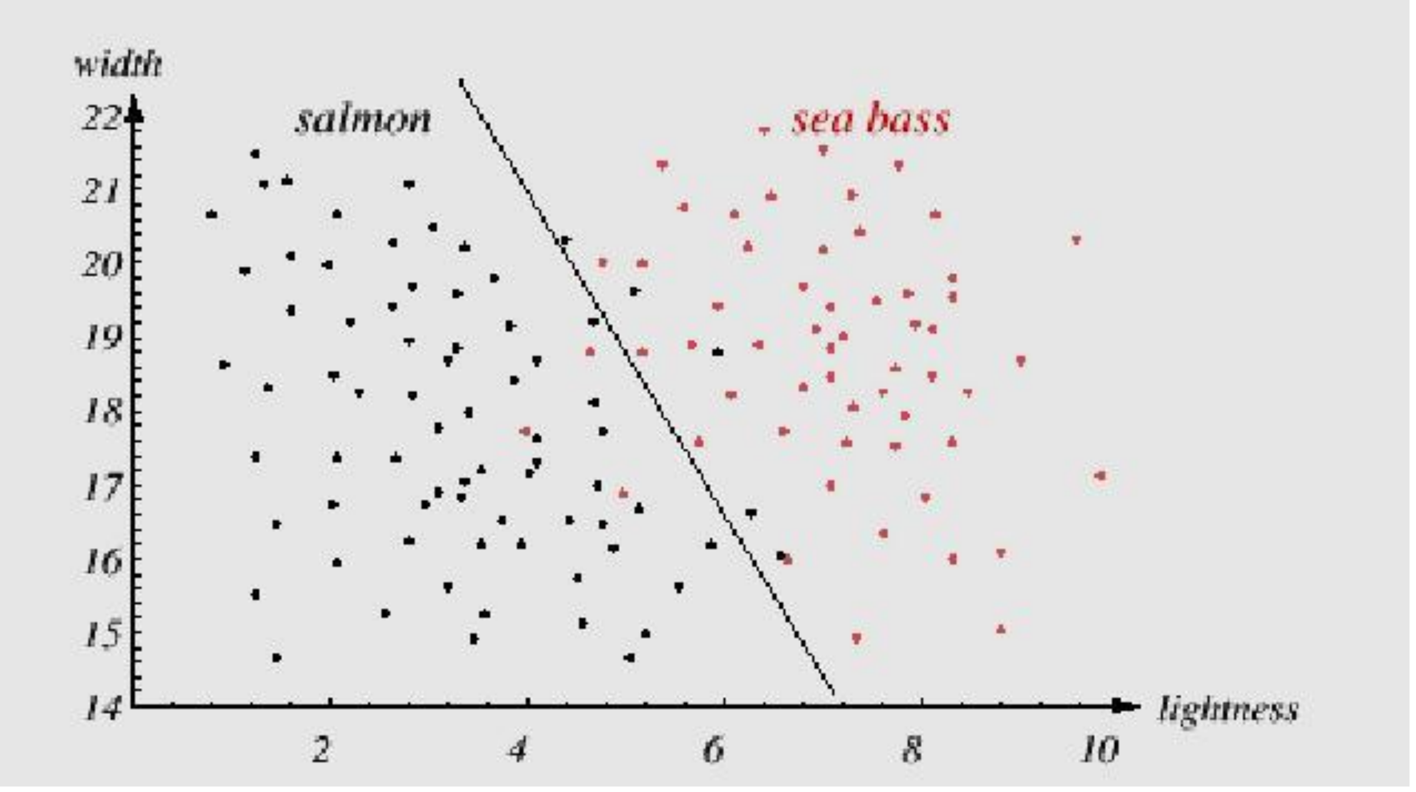
Selection Criterion and Decision Boundary

Adopt the lightness and add the width of the fish

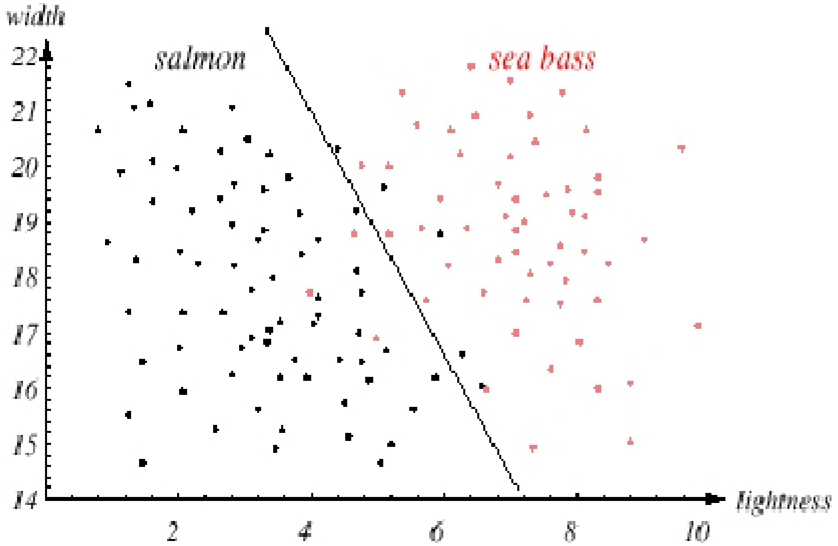
Fish → $x^T = [x_1, x_2]$

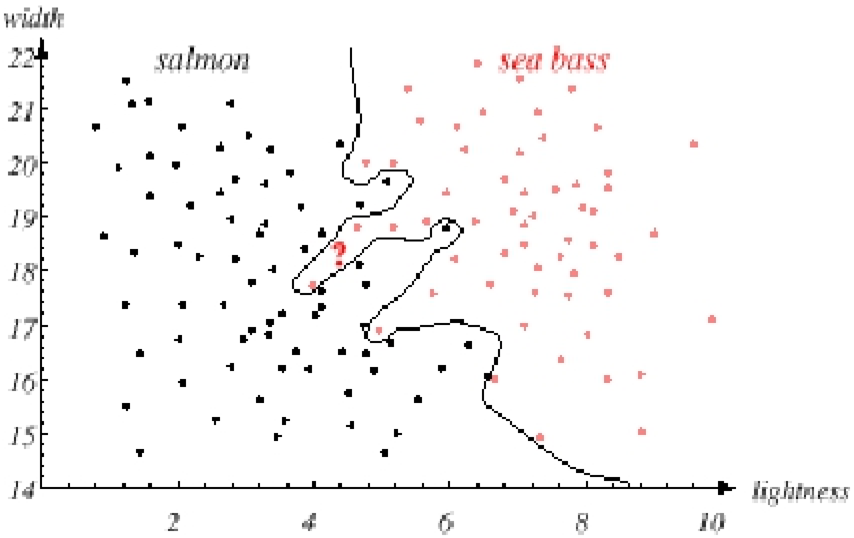
Lightness

Width



Decision Boundary Selection



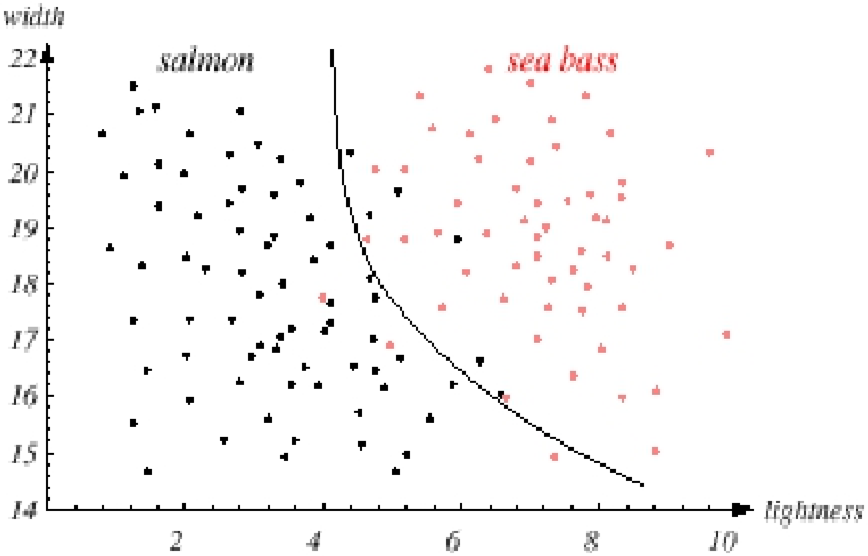


Which of the boundaries would you choose?

Simple linear boundary (training error > 0)

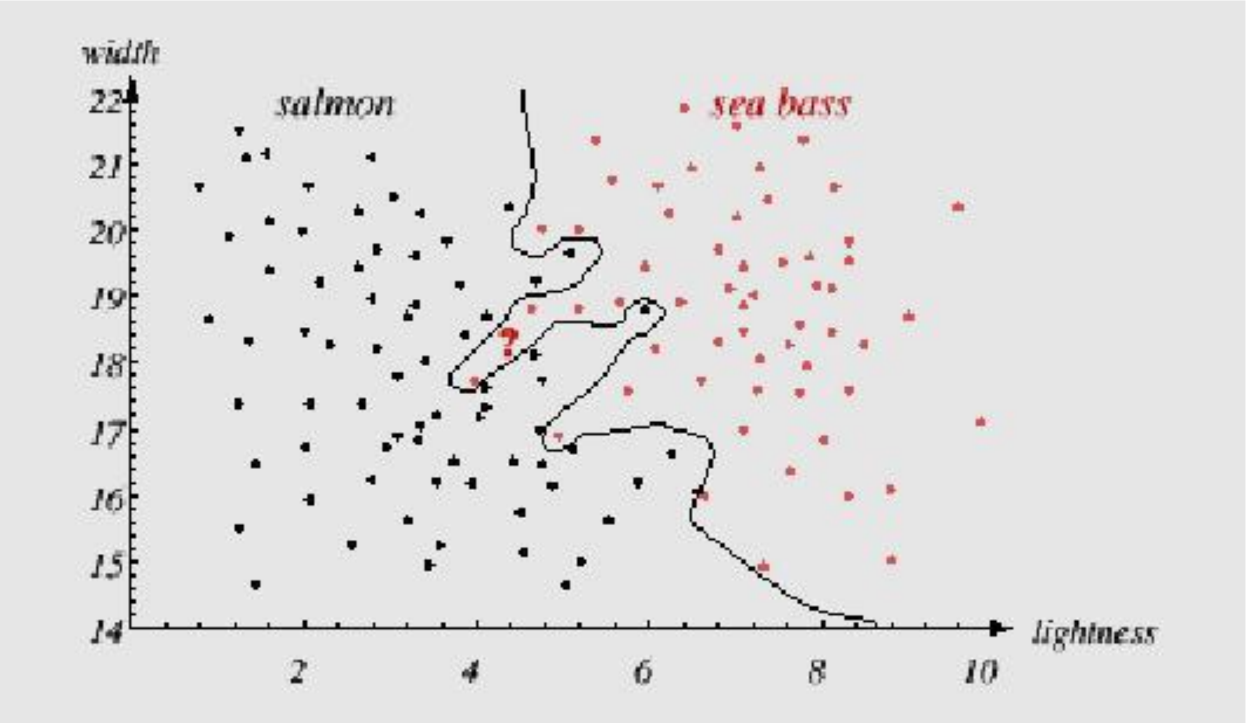
Nonlinear complex boundary (tr. error = 0)

Simpler nonlinear boundary (tr. error > 0)



Generalization and Decision Boundary

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

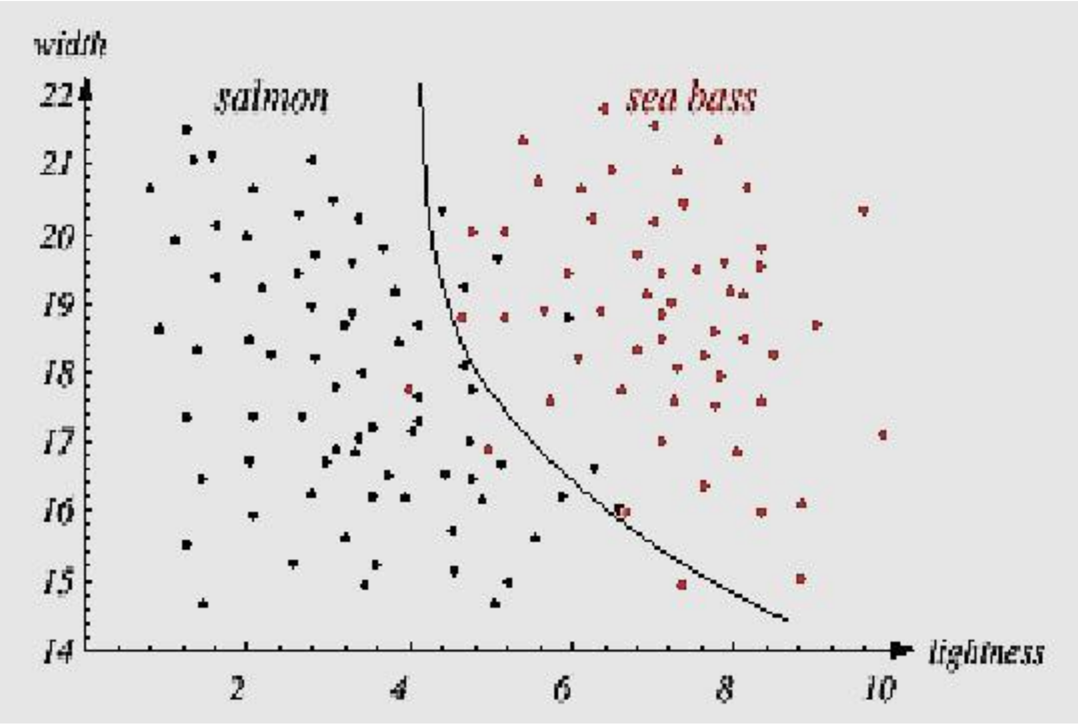


However, the central aim of designing a classifier is to correctly classify novel input (classify future samples)

Issue of generalization

Selected decision boundary

A simpler but generalized boundary in this case may be the simple nonlinear curve as shown below:



Further features can be added to enhance the performance

However redundancy in information may not be desirable

Adjustment of decision boundary

A classifier, intuitively, is designed to minimize classification error, the total number of instances (fish) classified incorrectly.

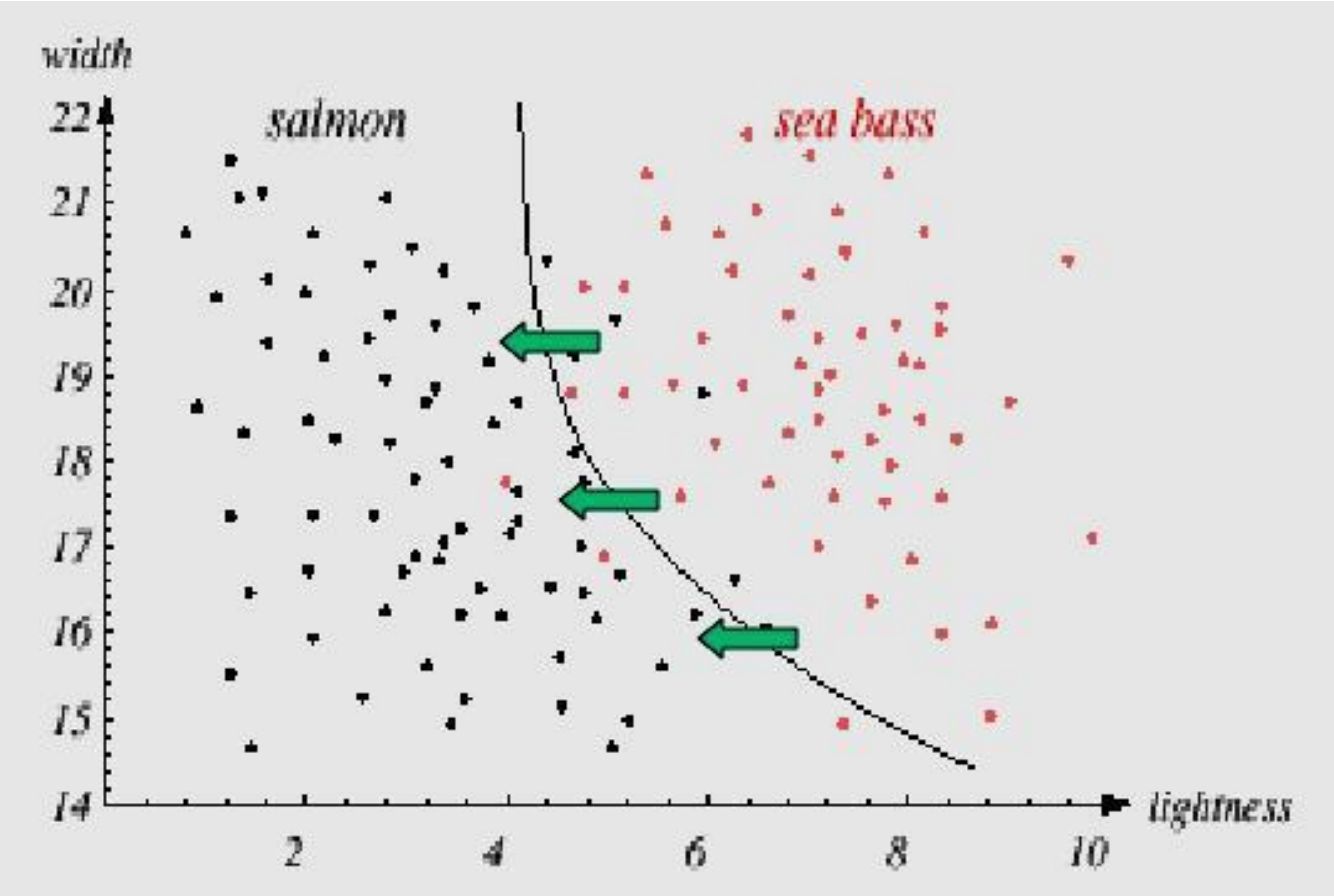
– Is this the best objective (cost) function to minimize? What kinds of error can be made? Are they all equally bad? What is the real cost of making an error?

  • *Sea bass misclassified as salmon: Pleasant surprise for the consumer, tastier fish*
  • *Salmon misclassified as sea bass: Customer upset, paid too much for inferior fish*

– We may want to adjust our decision boundary to minimize overall risk –in this case, second type error is more costly, so we may want to minimize this error.

Adjustment of decision boundary

Error is minimized by moving the threshold to the left



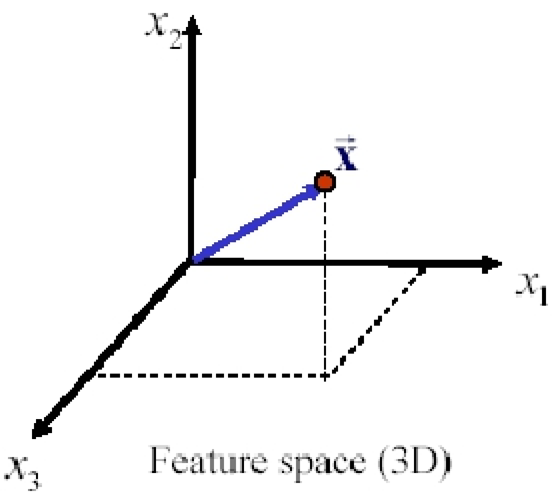
Terminologies in PR

Features: a set of variables believed to carry discriminating and characterizing information about the objects under consideration

Feature vector: A collection of *d* features, ordered in some meaningful way into a *d-dimensional* column vector, that represents the signature of the object to be identified.

Feature space: The *d-dimensional* space in which the feature vectors lie. A *d-dimensional* vector in a *d-dimensional* space constitutes a point in that space.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \textit{feature 1} \\ \textit{feature 2} \\ \\ \textit{feature d} \end{array}$$



Feature space (3D)

Terminologies in PR

Decision boundary:A boundary in the *d-dimensional* feature space that separates patterns of different classes from each other.

Training Data: Data used during training of a classifier for which the correct labels are known

Field Test Data: Unknown data to be classified –for measurements from which the classifier is trained. The correct class of this data are not known

Terminologies in PR

Cost Function: A quantitative measure that represents the cost of making an error. The classifier is trained to minimize this function.

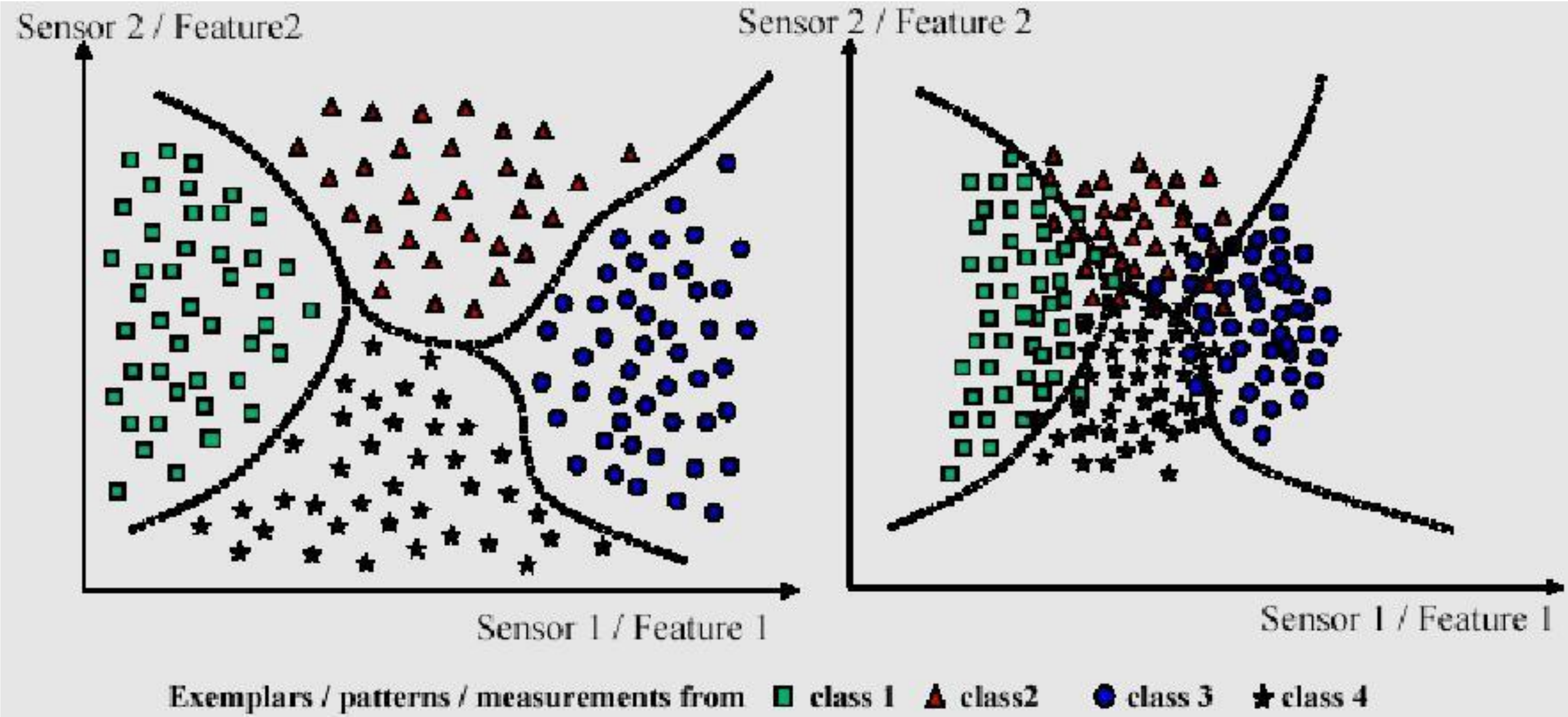
Classifier: An algorithm which adjusts its parameters to find the correct decision boundaries –through a learning algorithm using a training dataset –such that a cost function is minimized.

Error: Incorrect labeling of the data by the classifier

Cost of error: Cost of making a decision, in particular an incorrect one –not all errors are equally costly!

Training Performance: The ability / performance of the classifier in correctly identifying the classes of the training data, which it has already seen. It may not be a good indicator of the generalization performance.

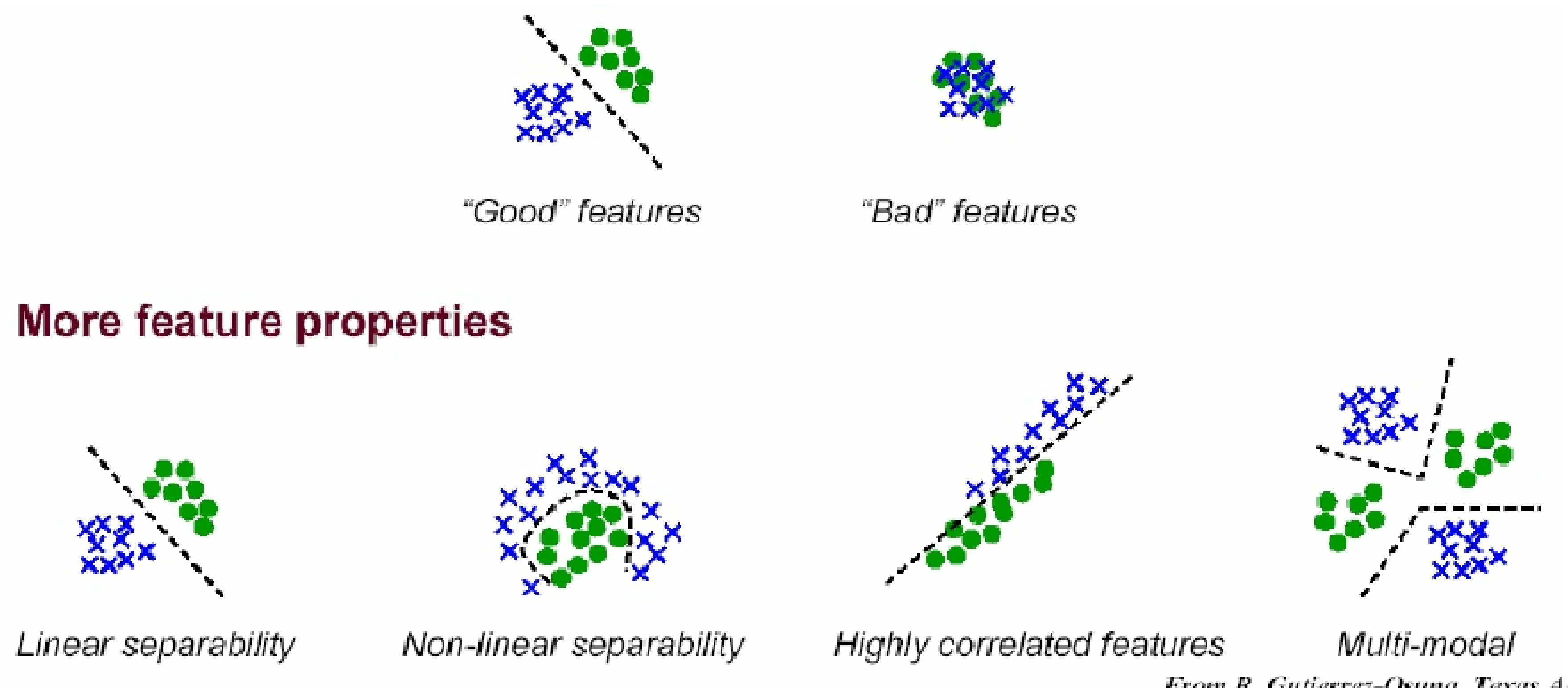
Cooperative vs. Uncooperative data



Good Features vs. Bad Features

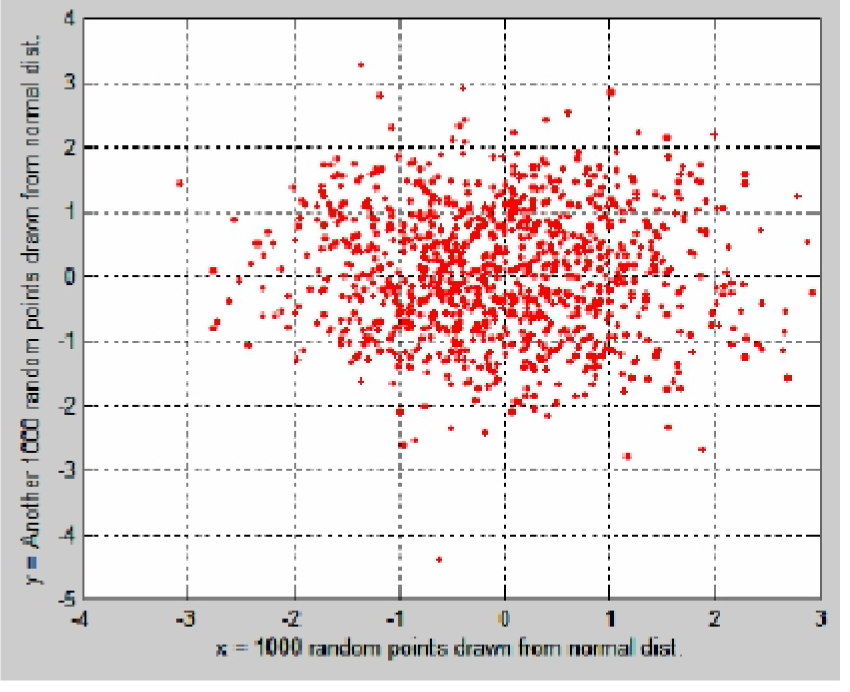
Ideally, for a given group of patterns coming from the same class, feature values should all be similar

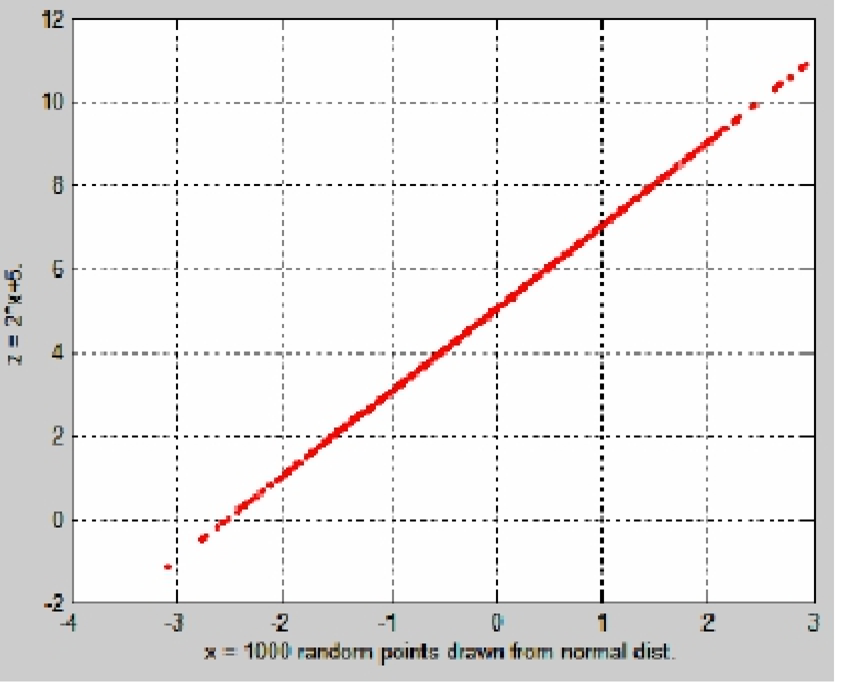
For patterns coming from different classes, the feature values should be different



From R. Gutierrez-Osuna, Texas A&M

Features Correlation





Components of Pattern Recognition System

What is the cost of making a decision –what is the risk of an incorrect decision?

Determine the right type of classifier, right type of training algorithm and right type of classifier parameters.

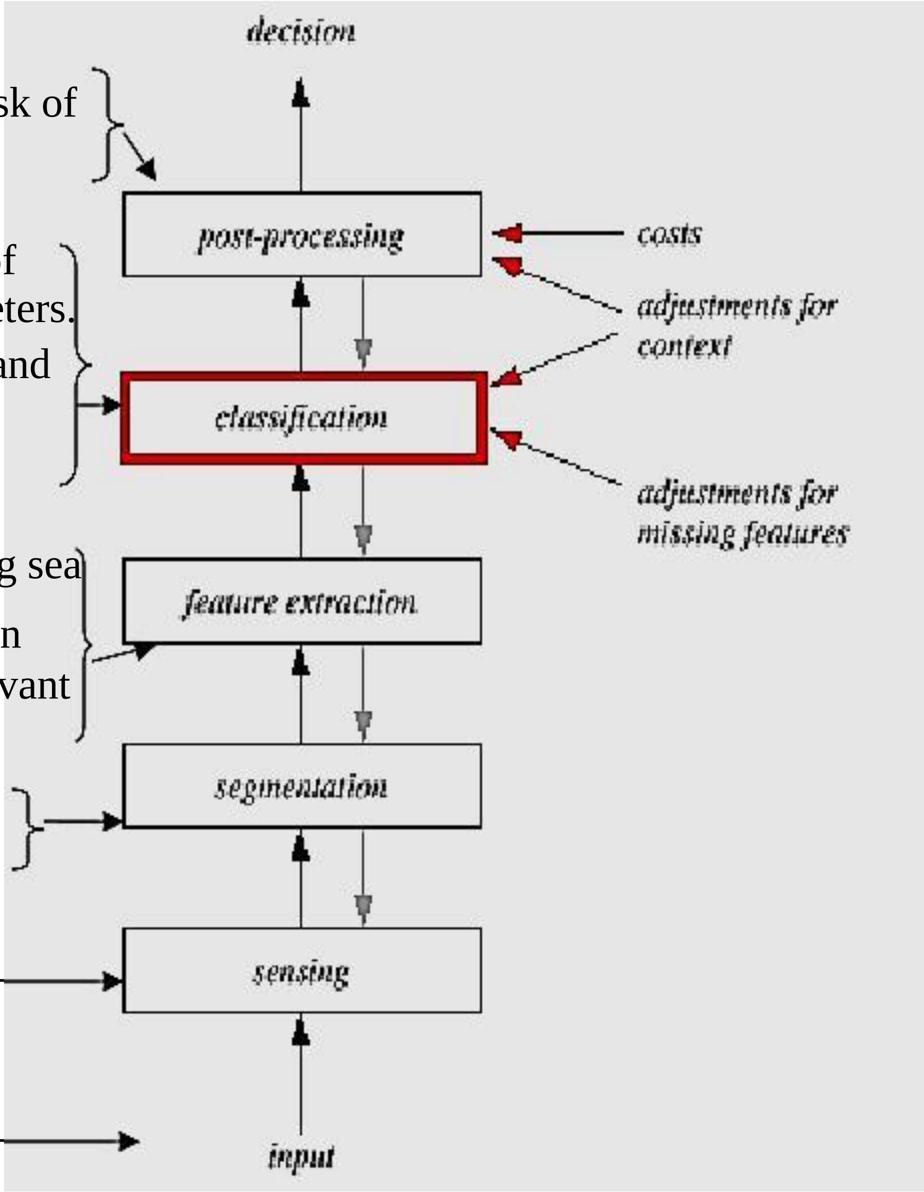
How to choose a classifier that minimizes the risk and maximizes the performance? Can we add prior knowledge, context to the classifier?

Identify and extract features relevant to distinguishing sea bass from salmon, but invariant to noise, occlusion rotations, scale, shift. What if we already have irrelevant features? Dimensionality reduction?

Separate & isolate fish from background image

CCD Camera on a conveyor belt

Fish

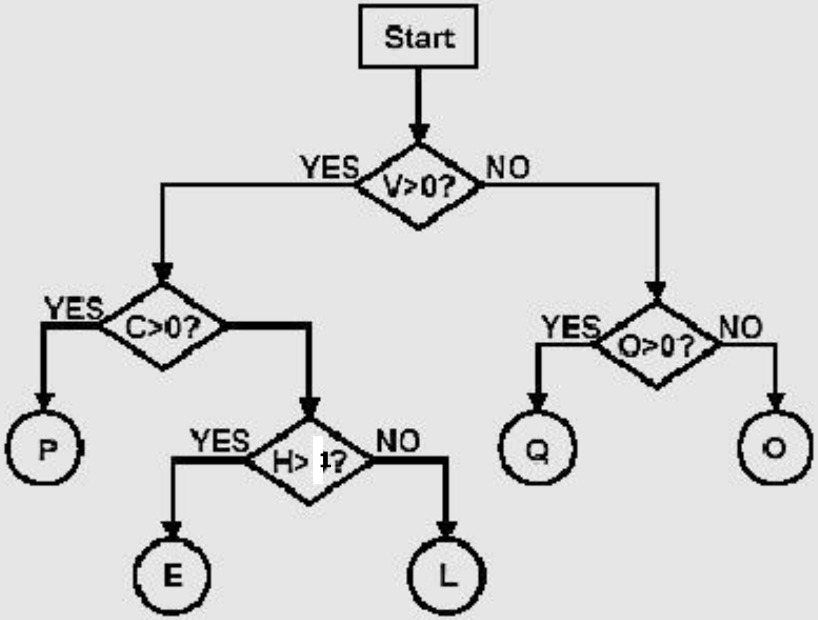


Example of PR

Consider the problem of recognizing the letters L,P,O,E,Q

- Determine a sufficient set of features
- Design a tree-structured classifier

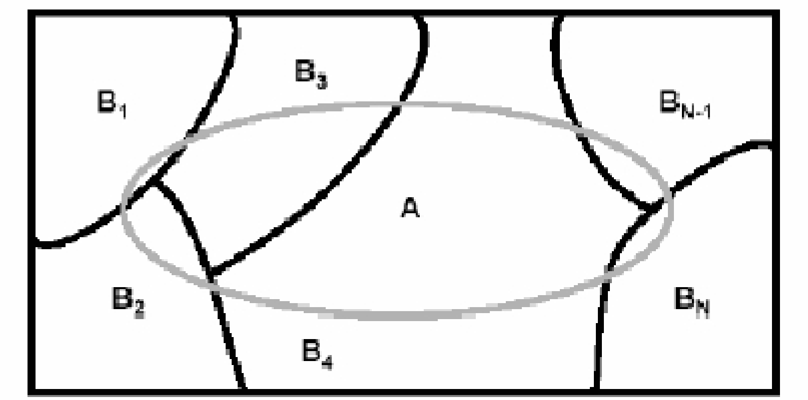
Character	Features			
	Vertical straight lines	*Horizontal straight lines*	*Oblique straight lines*	*Curved lines*
L	1	1	0	0
P	1	0	0	1
O	0	0	0	1
E	1	3	0	0
Q	0	0	1	1

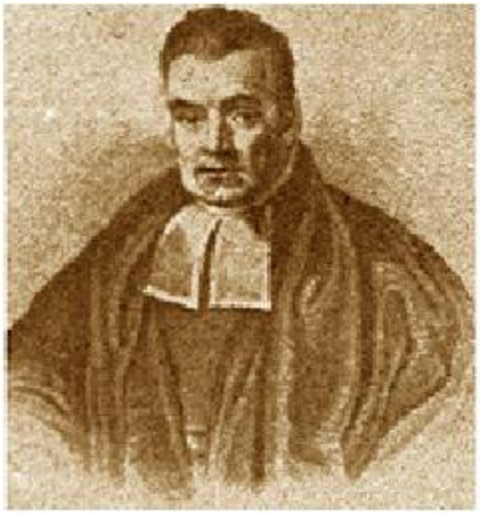


Bayesian Decision Theory

Bayes Rule

We now pose the following question: Given that the event *A* has occurred. What is the probability that any single one of the event *B*'s occur?




Rev. Thomas Bayes (1702-1761)

$$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{k=1}^{N} P(A \mid B_k) \cdot P(B_k)}$$

This is known as the Bayes rule

The Real World: The World of Many Dimensions

In most practical applications, we have more then one feature, and therefore the random variable x must be replaced with a random vector **x**. $P(x) \Rightarrow P(\mathbf{x})$

The joint probability mass function $P(\mathbf{x})$ still satisfies the axioms of probability

The Bayes rule is then

$$\boxed{P(\omega_j \mid \mathbf{x}) = \frac{P\left(\mathbf{x} \cap \omega_j\right)}{P(\mathbf{x})} = \frac{P(\mathbf{x} \mid \omega_j) \cdot P(\omega_j)}{\sum_{k=1}^{C} P(\mathbf{x} \mid \omega_k) \cdot P(\omega_k)}}$$

While the notation changes only slightly, the implications are quite substantial

Bayes Decision theory

Based on quantifying the trade offs betweens various classification decisions using a probabilistic approach

The theory assumes:

- Decision problem can be posed in probabilistic terms
- All relevant probability values are known (in practice this is not true)

The sea bass/salmon example

State of nature: ω_1: Sea bass, ω_2: Salmon

Assume that we know the probabilities of observing sea bass and salmons, $P(\omega_1)$ and $P(\omega_2)$ (*prior* probabilities), for a particular location of fishing and time of year

Based on this information, how would you guess that the type of the next fish to be caught?

- *Decide* ω_1 *if* $P(\omega_1) > P(\omega_2)$ *otherwise decide* ω_2

  A reasonable decision rule?

If the catch of salmon and sea bass is equi-probable

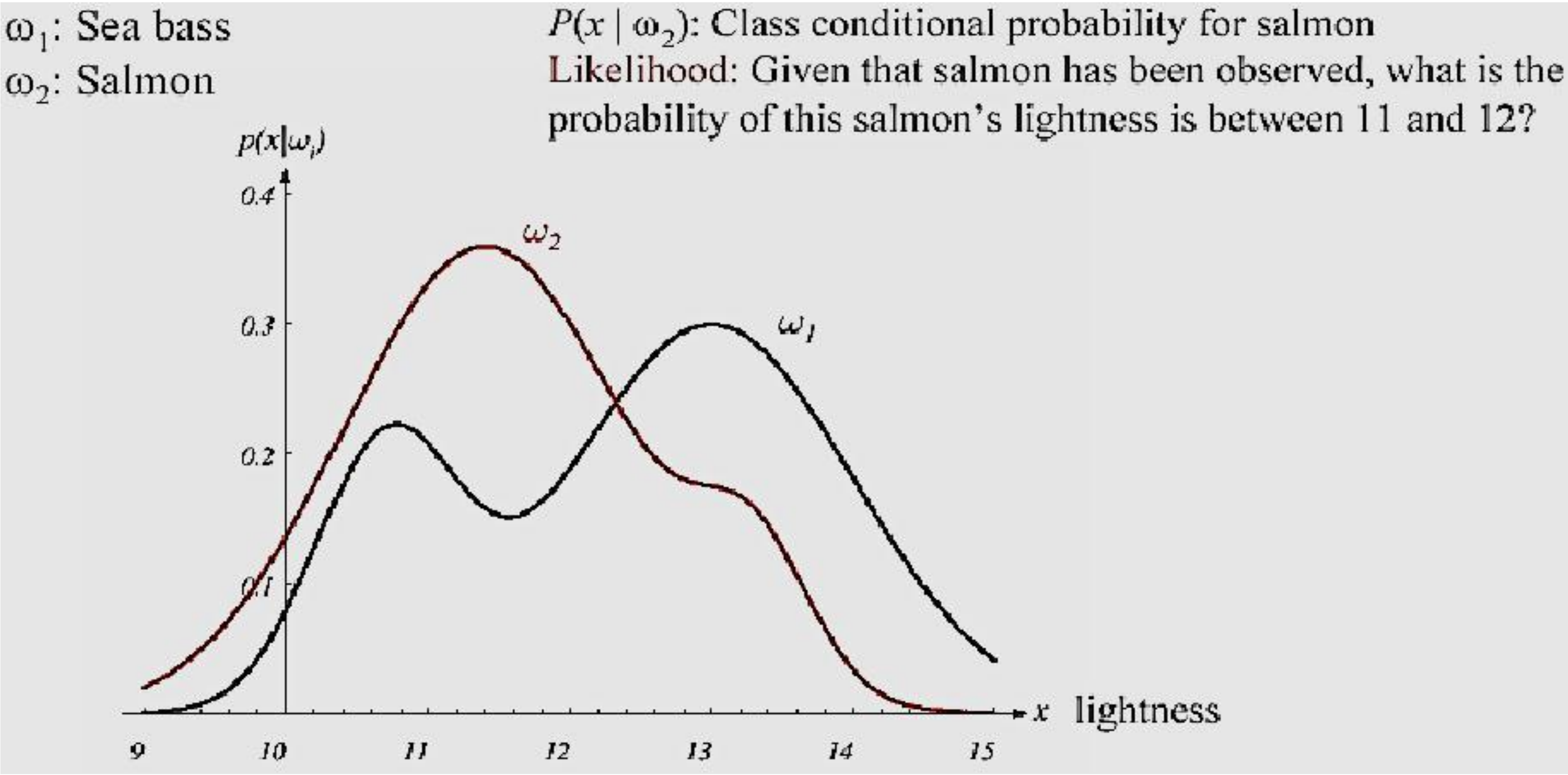
$P(\omega_1) = P(\omega_2)$ (uniform priors)

$P(\omega_1) + P(\omega_2) = 1$ (exclusivity and exhaustivity)

This rule will not be applicable

Class Conditional Probabilities

$p(x|\omega_1)$ and $p(x|\omega_2)$ describe the difference in lightness between populations of sea bass and salmon



Posterior Probability

Suppose, we know $P(\omega_1)$, $P(\omega_2)$, $p(x|\omega_1)$ and $p(x|\omega_2)$ and that we have observed the value of the feature (a random variable) x

- How would you decide on the "state of nature"–type of fish, based on this info?
- Bayes theory allows us to compute the posterior probabilities from prior and class-conditional probabilities

Likelihood: The (class-conditional) probability of observing a feature value of x, given that the correct class is ω_j. All things being equal, the category with higher class conditional probability is more "likely"to be the correct class.

Prior Probability: The total probability of correct class being class ω_j determined based on prior experience

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) \cdot P(\omega_j)}{\sum_j p(x \mid \omega_j) \cdot P(\omega_j)} = \frac{p(x \mid \omega_j) \cdot P(\omega_j)}{p(x)}$$

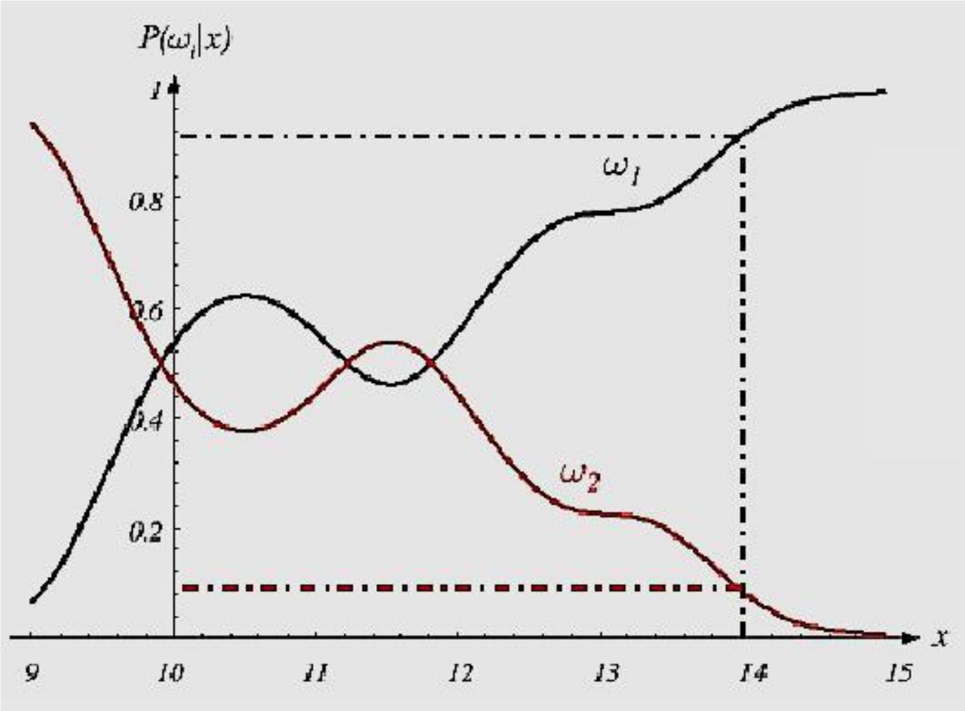
Posterior Probability: The (conditional) probability of correct class being ùj, given that feature value x has been observed

Evidence: The total probability of observing the feature value as x

Posterior Probabilities

Bayes rule allows us to compute the posterior probability (difficult to determine) from prior probabilities, likelihood and the evidence (easier to determine).

Posterior probabilities for priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$. For example, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0.



Which class would you choose now?

What is probability of making an error with this decision?

Bayes Decision Rule

Decision given the posterior probabilities

X is an observation for which:
if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⇨ True state of nature = ω_1
if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⇨ True state of nature = ω_2

That is Choose the class that has the larger posterior probability !

If there are multiple features, $\mathbf{x}=\{x_1, x_2,\ldots, x_d\}$ and multiple classes

Choose ω_i if $P(\omega_i \mid \mathbf{x}) > P(\omega_j \mid \mathbf{x})$ *for all* $i = 1, 2, \ldots, c$

Error:

whenever we observe a particular x, the probability of error is :

$$P(error \mid x) = P(\omega_1 \mid x) \text{ if we decide } \omega_2$$

$$P(error \mid x) = P(\omega_2 \mid x) \text{ if we decide } \omega_1$$

Therefore:

$$P(error \mid x) = \min [P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

(Bayes decision)

Bayesian Decision Theory – Generalization

Generalization of the preceding ideas

- Use of more than one feature
- Use more than two states of nature
- Allowing other actions and not only decide on the state of nature
- Unequal error cost

Allowing actions other than classification primarily allows the possibility of rejection: refusing to make a decision in close or bad cases!

The loss function states how costly each action taken is

Two-category classification

Definitions

α_1 : deciding ω_1

α_2 : deciding ω_2

$\varepsilon_{ij} = \varepsilon(\alpha_i \mid \omega_j)$, loss incurred for deciding ω_i when the true state of nature is ω_j

Conditional risk:

$$R(\alpha_1 \mid x) = \varepsilon_{11}P(\omega_1 \mid x) + \varepsilon_{12}P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \varepsilon_{21}P(\omega_1 \mid x) + \varepsilon_{22}P(\omega_2 \mid x)$$

Our rule is the following:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$ action α_1: "decide ω_1" is taken

End

Intro to Pattern Recognition