PATTERN RECOGNITION

Lecture 2: Basics of Pattern Classification

Professor Aly A. Farag

Computer Vision and Image Processing Laboratory

University of Louisville

URL: www.cvip.uofl.edu ; E-mail: aly.farag@louisville.edu

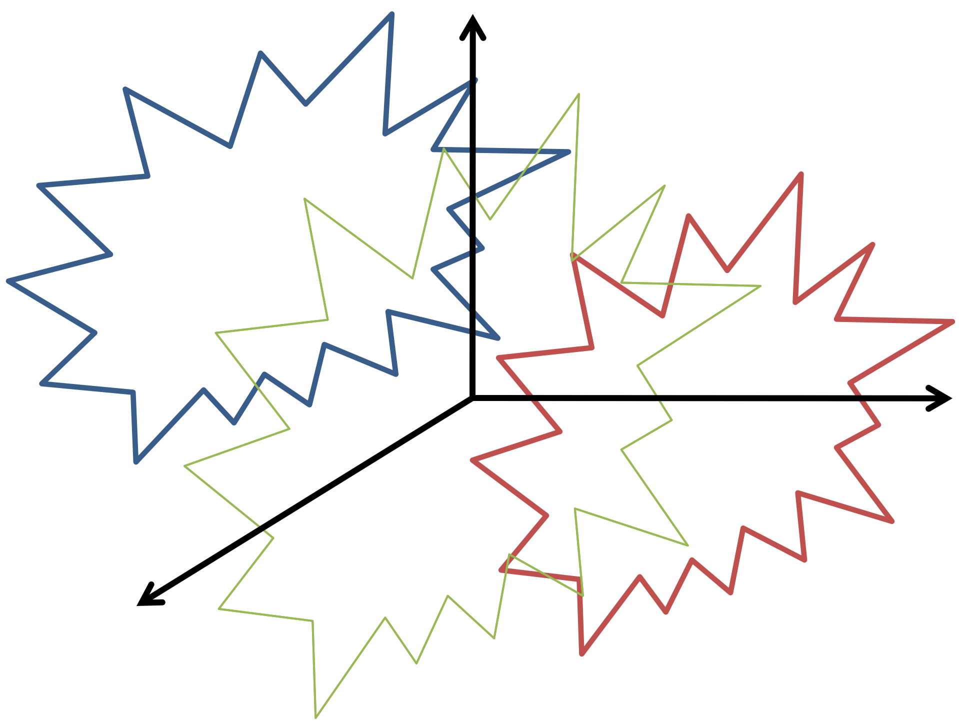
Planned for ECE 620 and ECE 655 - Summer 2011

TA/Grader: Melih Aslan; CVIP Lab Rm 6, msaslan01@lousiville.edu

Pattern Classification

Goal: Given measurements (outcomes) of experiments that share common attributes, how to separate these measurements?

Illustration:



An Example: Sorting fish!

Real world approach!!!

No, let's be Americanized!

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

Example:

Sea bass

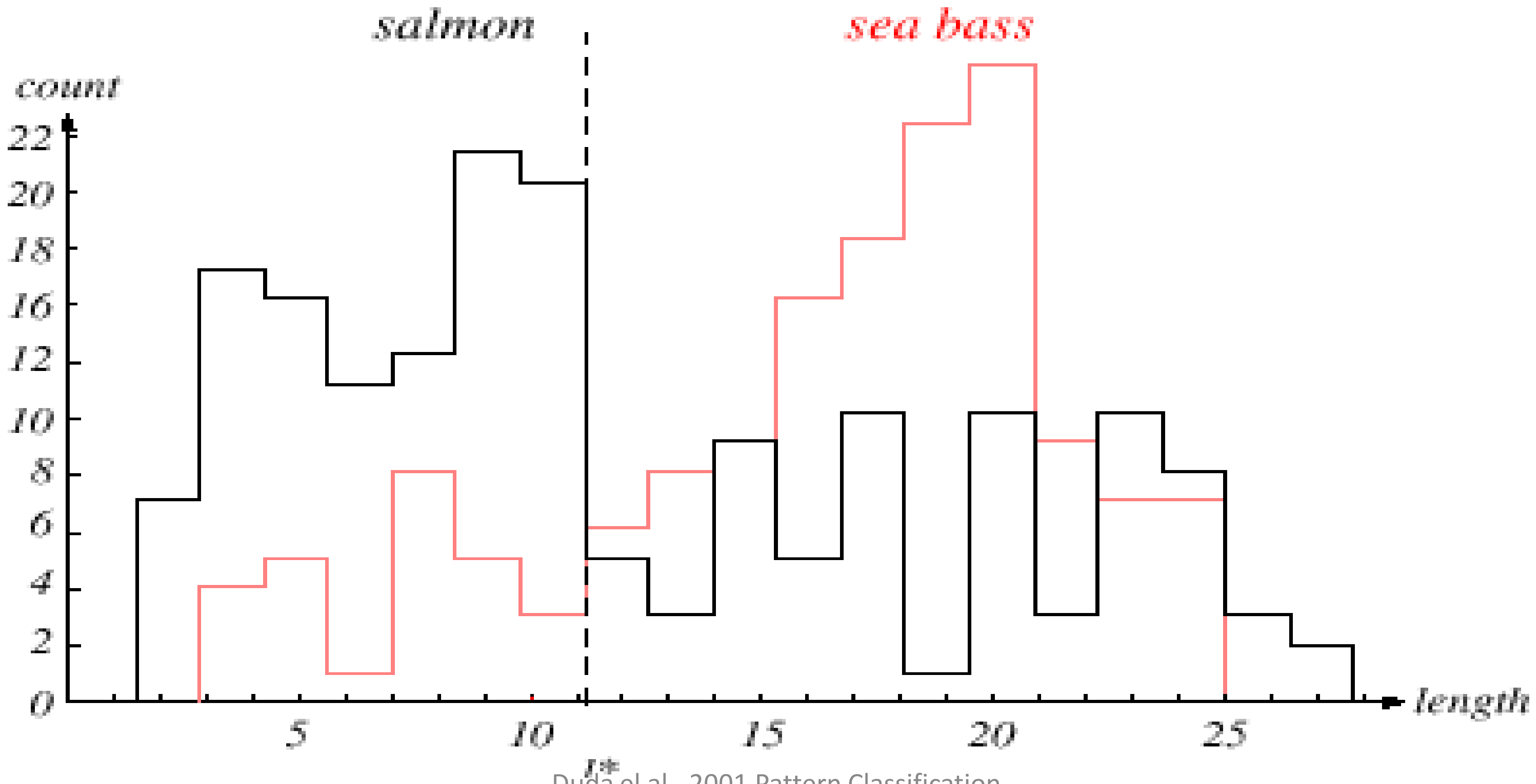
Species

Salmon

Duda el al., 2001 Pattern Classification, Chap. 1

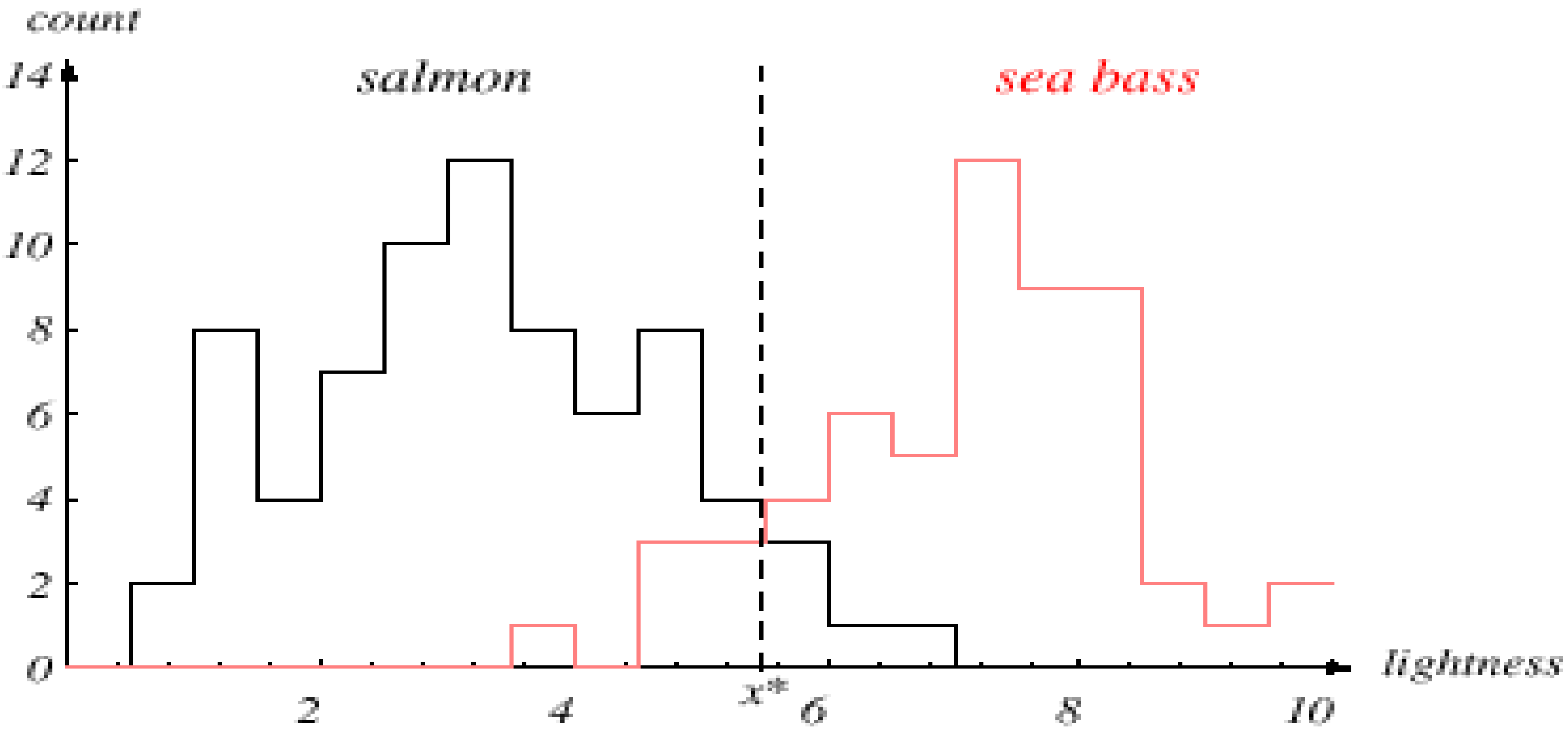
Classification

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



The length is a poor feature alone!

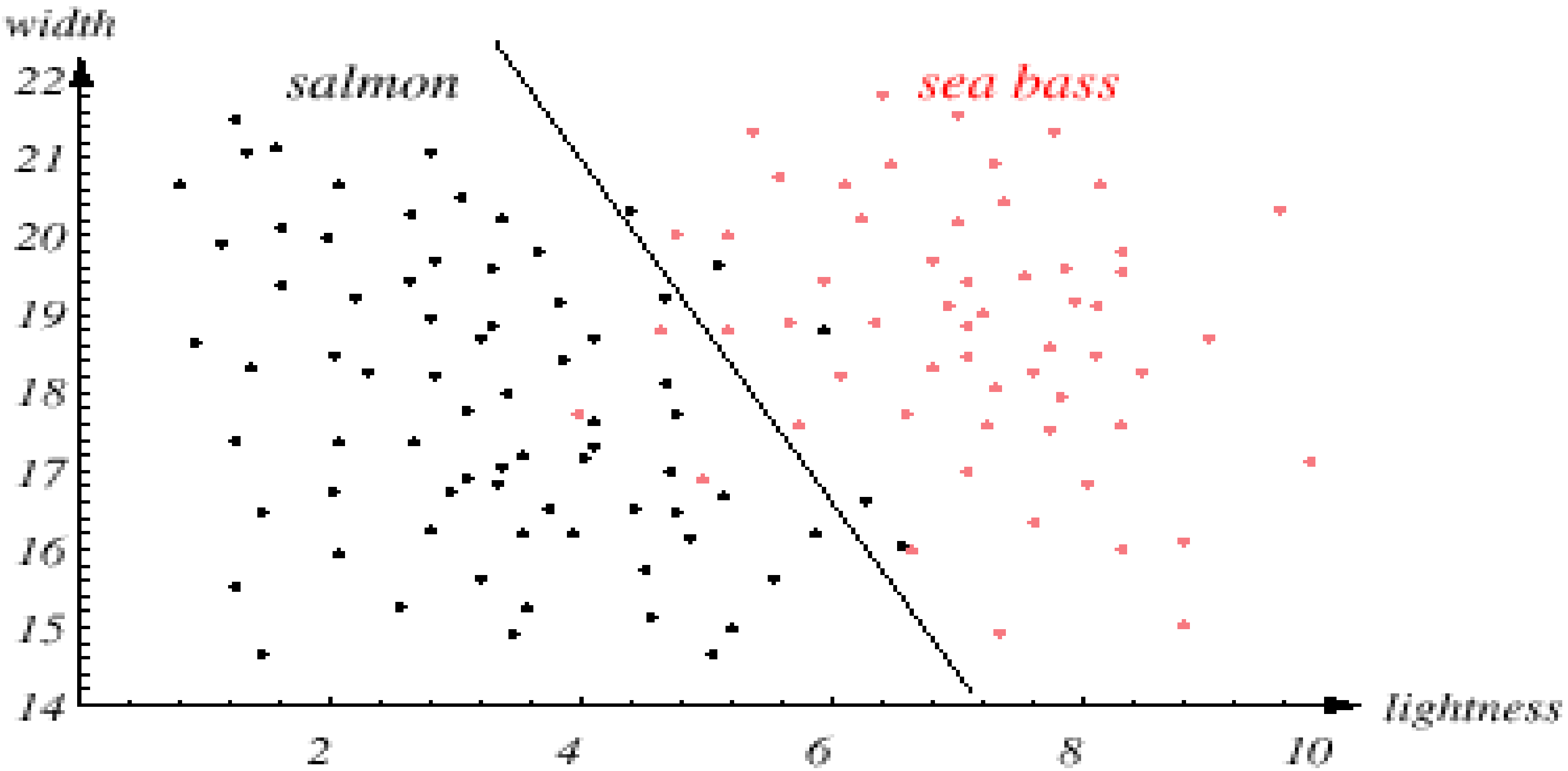
→ Select the lightness as a possible feature.



→ Adopt the lightness and add the width of the fish

$x^T = [x_1, x_2]$

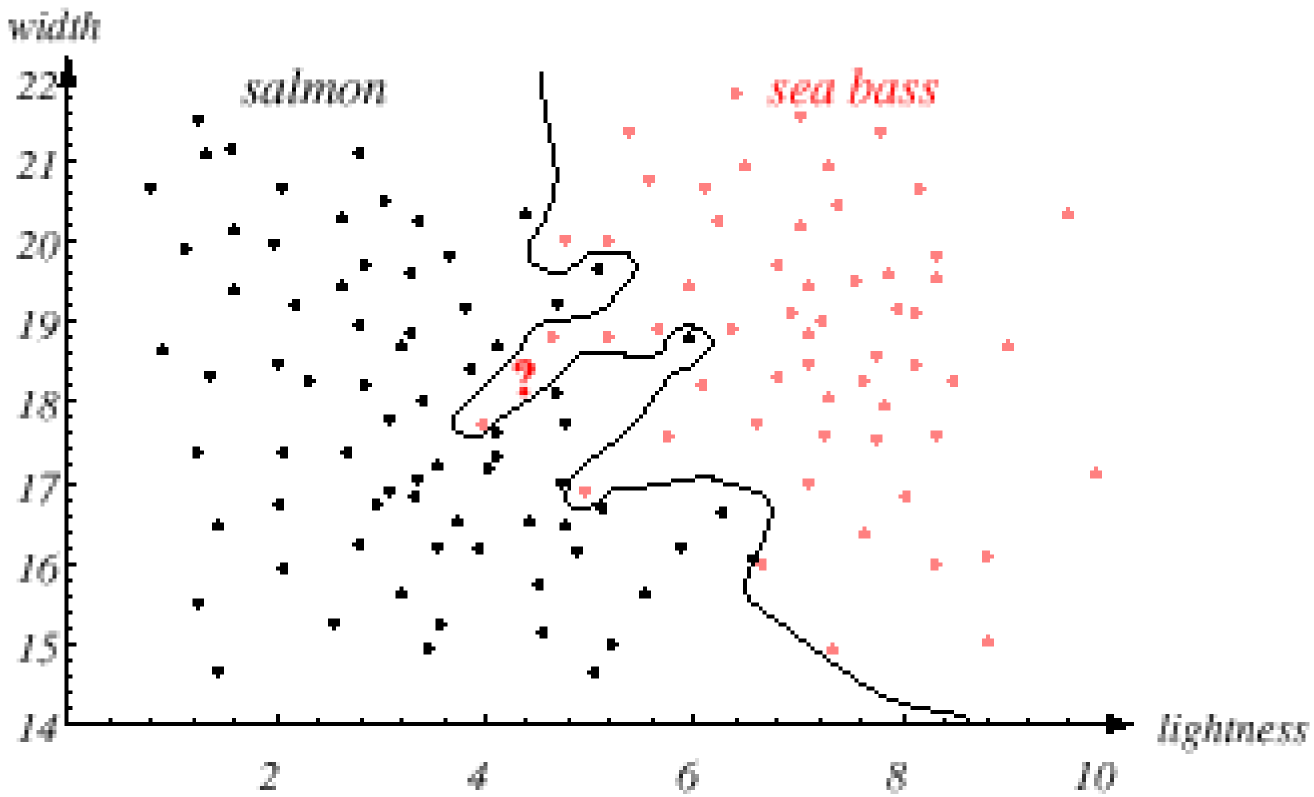
lightness width



Duda el al., 2001 Pattern Classification, Chap. 1

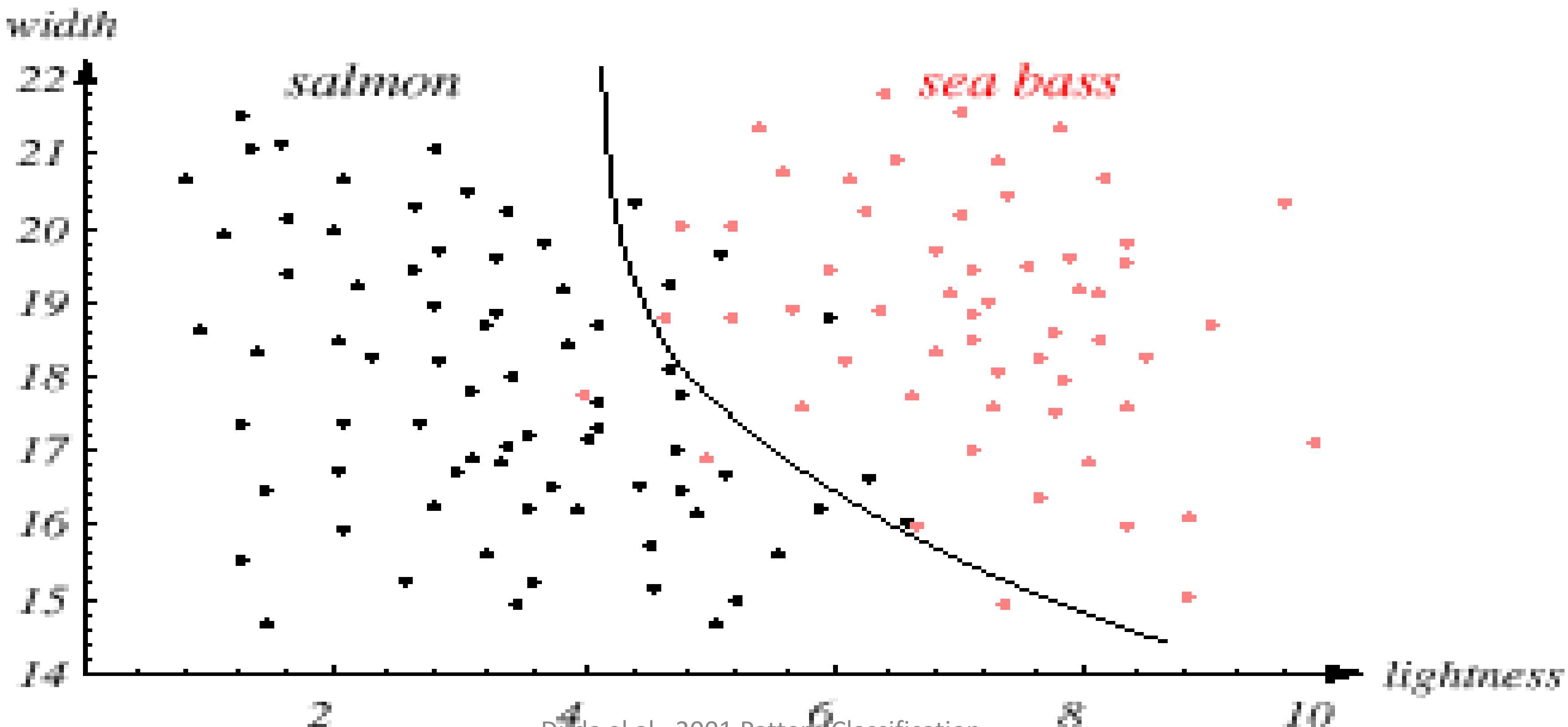
- We might add other features that are not correlated with the ones we already have.

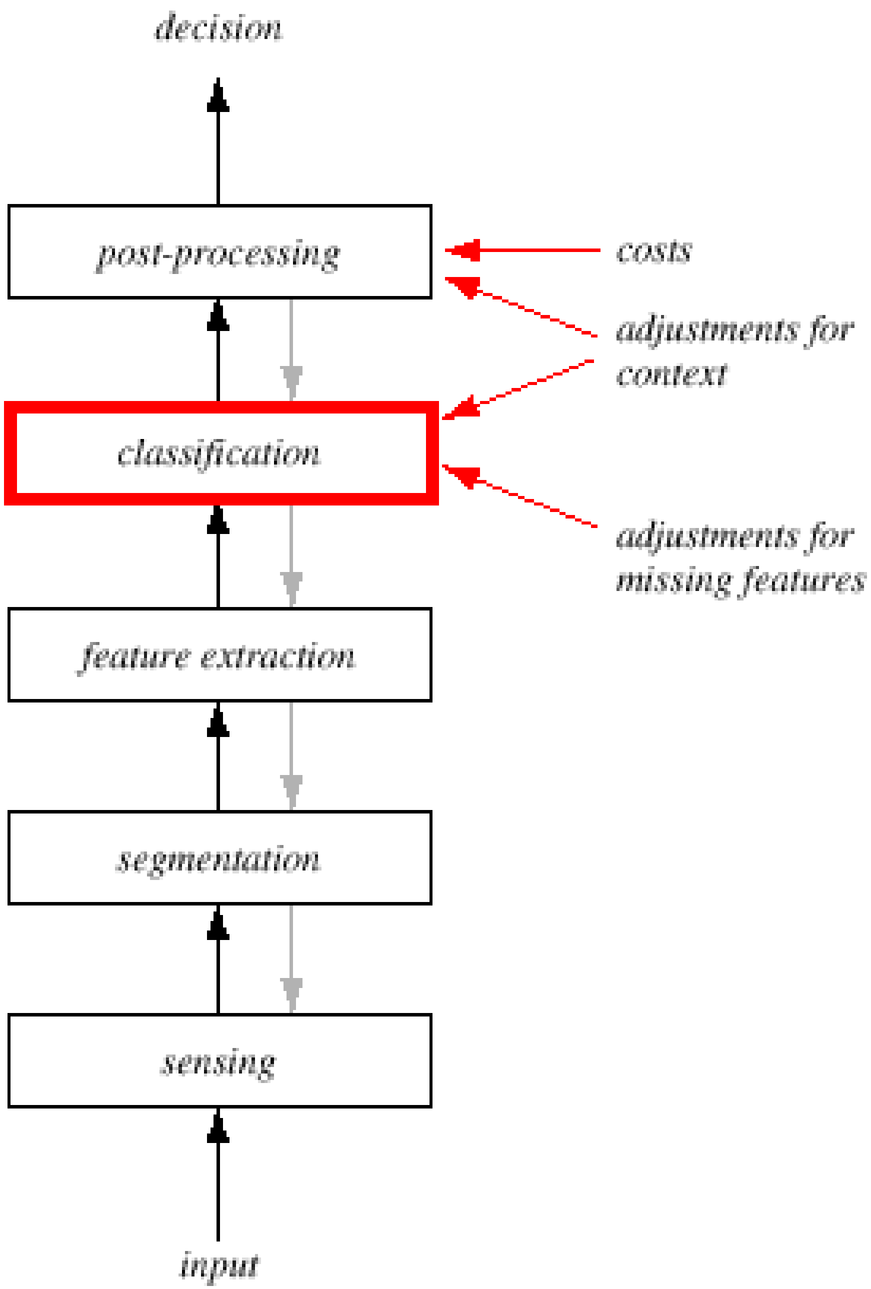
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:


width
salmon
sea bass
?
22
21
20
19
18
17
16
15
14
2
4
6
8
10
lightness

However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input

Issue of generalization!





Duda el al., 2001 Pattern Classification, Chap. 1

- **Feature extraction**
  - Discriminative features
  - Invariant features with respect to translation, rotation and scale.
- **Classification**
  - Use a feature vector provided by a feature extractor to assign the object to a category
- **Post Processing**
  - Exploit *context* input dependent information other than from the target pattern itself to improve performance

The Design Cycle

- Data collection
- Feature Choice
- Model Choice
- Training
- Evaluation
- Computational Complexity

Bayesian Decision Theory

Consider a binary (two-class) classification problem.

Decision rule with only the prior information:

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

Decision rule using of the class –conditional information:

– Decide ω_1 if $P(x \mid \omega_1) > P(x \mid \omega_2)$

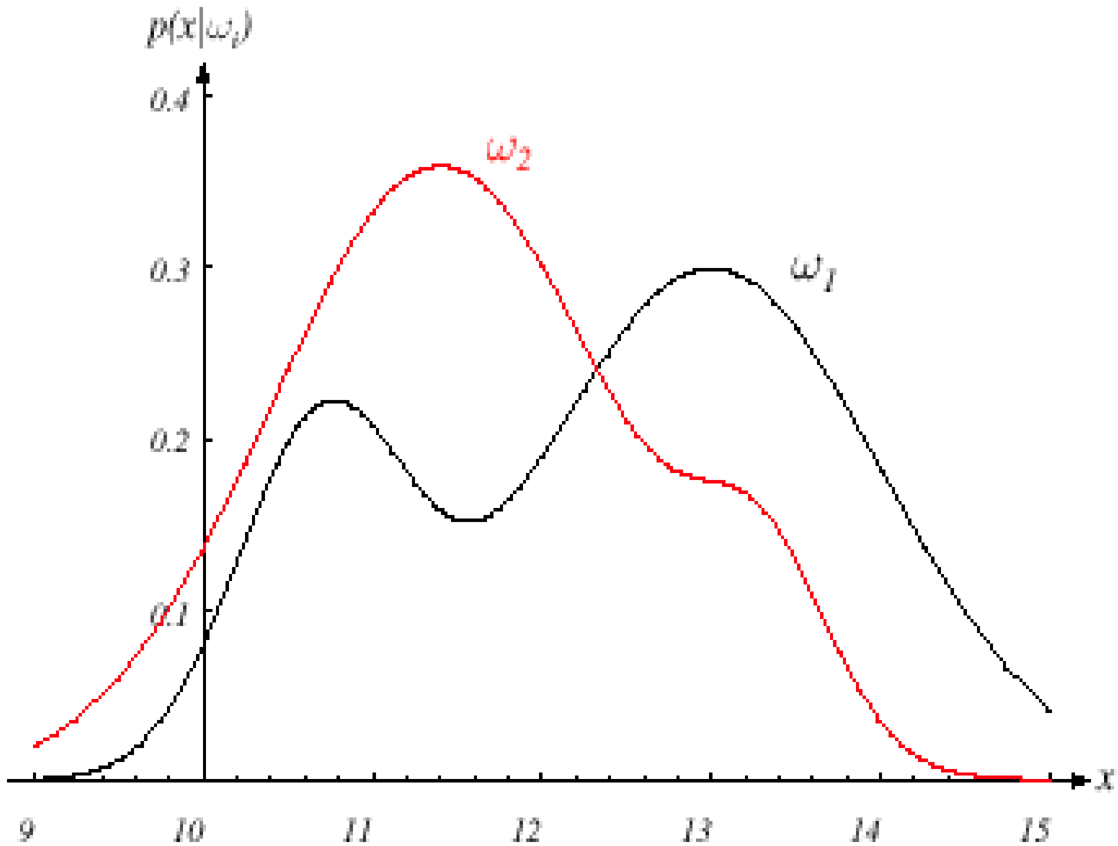


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Posterior, likelihood, evidence

$$P(\omega_j \mid x) = P(x \mid \omega_j) \,.\, P(\omega_j) / P(x)$$

– Where in case of two categories

$$P(x) = \sum_{j=1}^{j=2} P(x \mid \omega_j) P(\omega_j)$$

– Posterior = (Likelihood. Prior) / Evidence

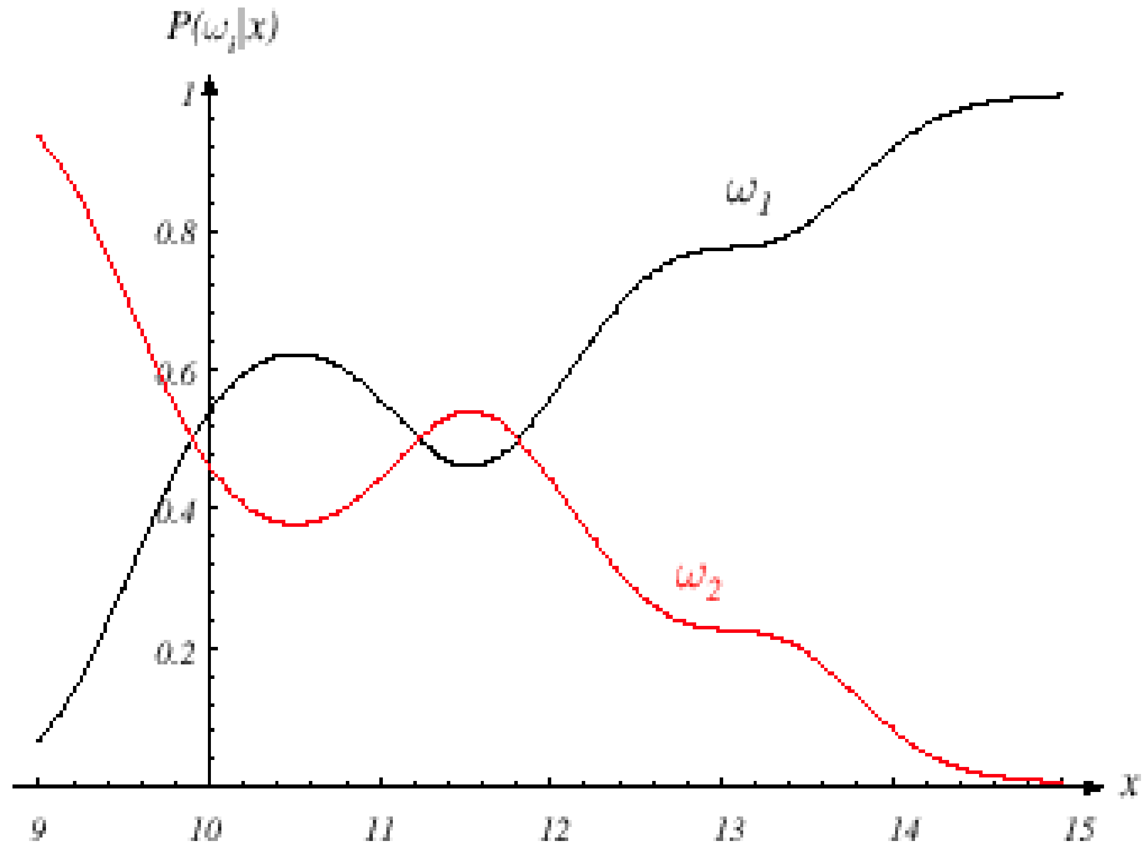


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Two-class (Binary) Classification:

- Decision given the posterior probabilities

X is an observation for which:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ $\Rightarrow$ True state of nature = ω_1
if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ $\Rightarrow$ True state of nature = ω_2

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

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

- Minimizing the probability of error

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

  Therefore:

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

(Bayes decision)

Bayesian Decision Theory: Minimum Risk Concept

Let $\{\omega_1, \omega_2, \ldots, \omega_c\}$ be the set of *c* states of nature (or "categories")

Let $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ be the set of possible actions

Let $\lambda(\alpha_i / \omega_j)$ be the loss incurred for taking

action α_i when the state of nature is ω_j

Overall risk

$$\underbrace{R = \text{Sum of all } R(\alpha_i \mid x)}_{\text{Conditional risk}} \text{ for } i = 1,\ldots,a$$

Minimizing R $\Longleftrightarrow$ Minimizing $R(\alpha_i \mid x)$ *for* $i = 1,\ldots, a$

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

for i = 1,…,a

Select the action α_i for which $R(\alpha_i \mid x)$ is minimum

R is minimum and R in this case is called the Bayes risk = best performance that can be achieved!

Example: Two-category classification

α_1 : deciding ω_1

α_2 : deciding ω_2

$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$

loss incurred for deciding ω_i when the true state of nature is ω_j

Conditional risk:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$

$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

Our rule is the following:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

action α_1: "decide ω_1" is taken

This results in the equivalent rule :

decide ω_1 if:

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) >$$
$$(\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

and decide ω_2 otherwise

Likelihood ratio:

The preceding rule is equivalent to the following rule:

$$\textbf{If} \quad \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Then take action α_1 (decide ω_1)

Otherwise take action α_2 (decide ω_2)

Optimal decision property:

"If the likelihood ratio exceeds a threshold value independent of the input pattern x, we can take optimal actions"

Exercise

Select the optimal decision where:

$\Omega = \{\omega_1, \omega_2\}$

$P(x \mid \omega_1) \Rightarrow$ N(2, 0.5) (Normal distribution)

$P(x \mid \omega_2) \Rightarrow$ N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

$$\lambda = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$