KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

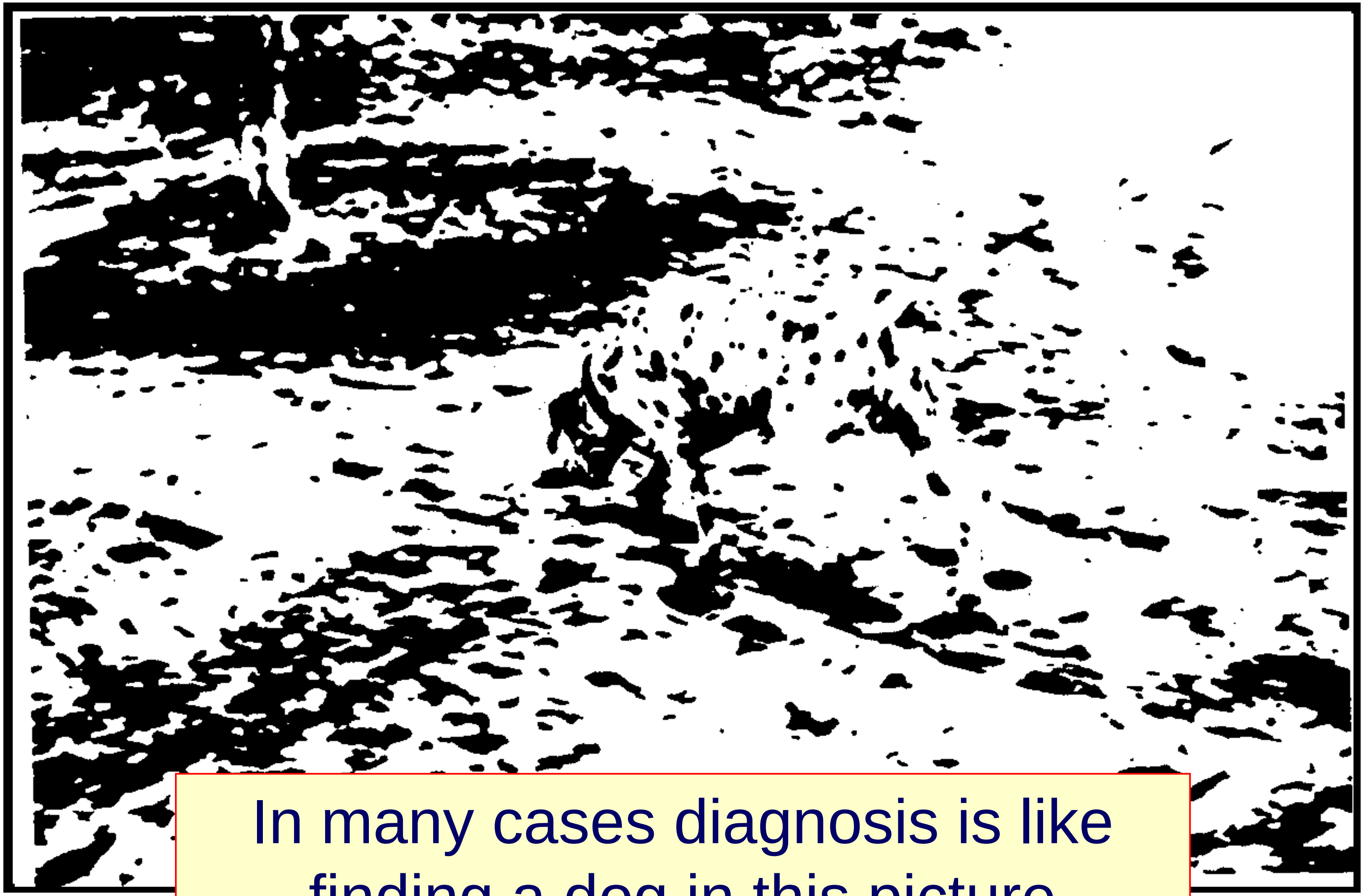
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

„Signal processing”

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego w projekcie pt.

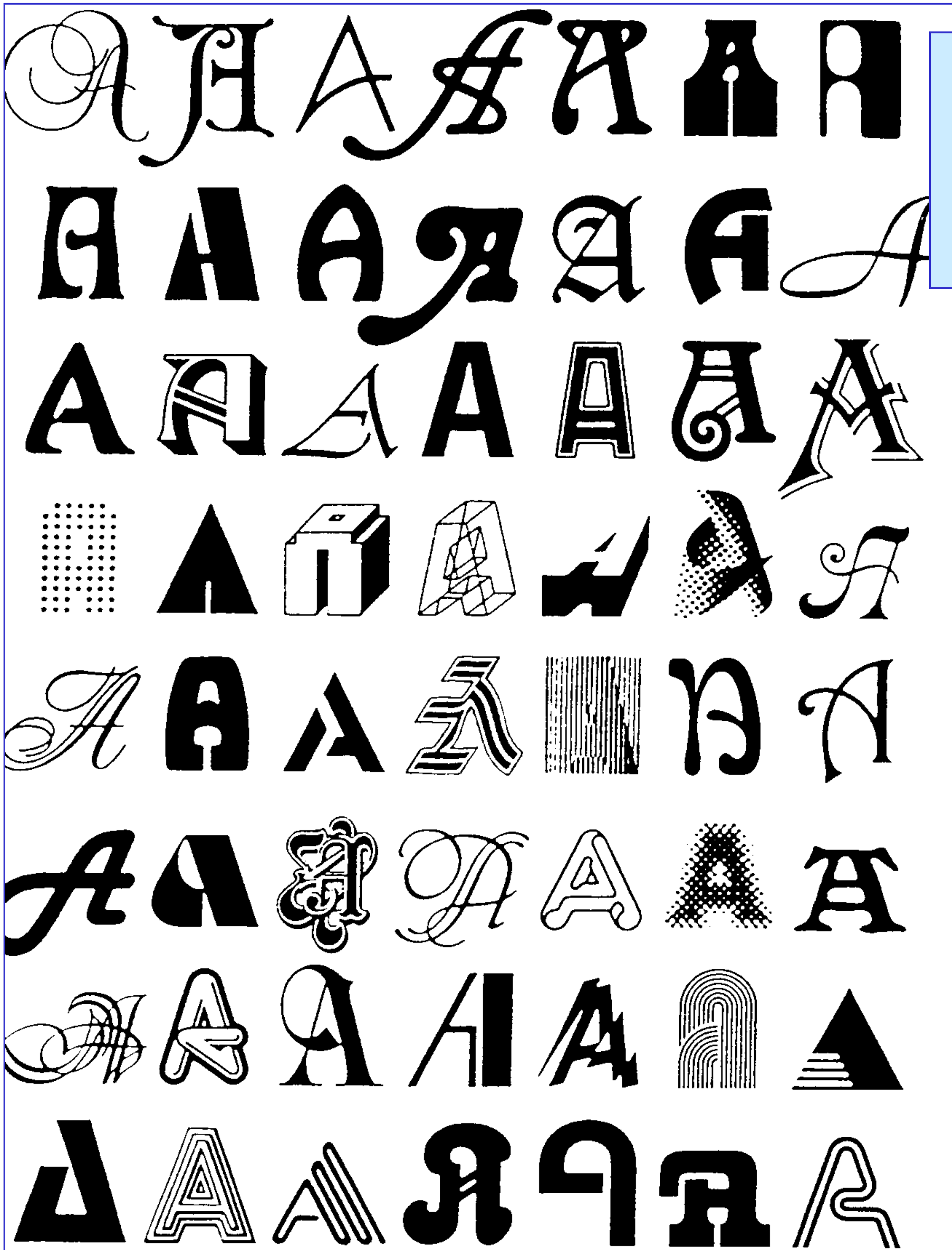
„Innowacyjna dydaktyka bez ograniczeń - zintegrowany rozwój Politechniki Łódzkiej - zarządzanie Uczelnią, nowoczesna oferta edukacyjna i wzmacniania zdolności do zatrudniania osób niepełnosprawnych”

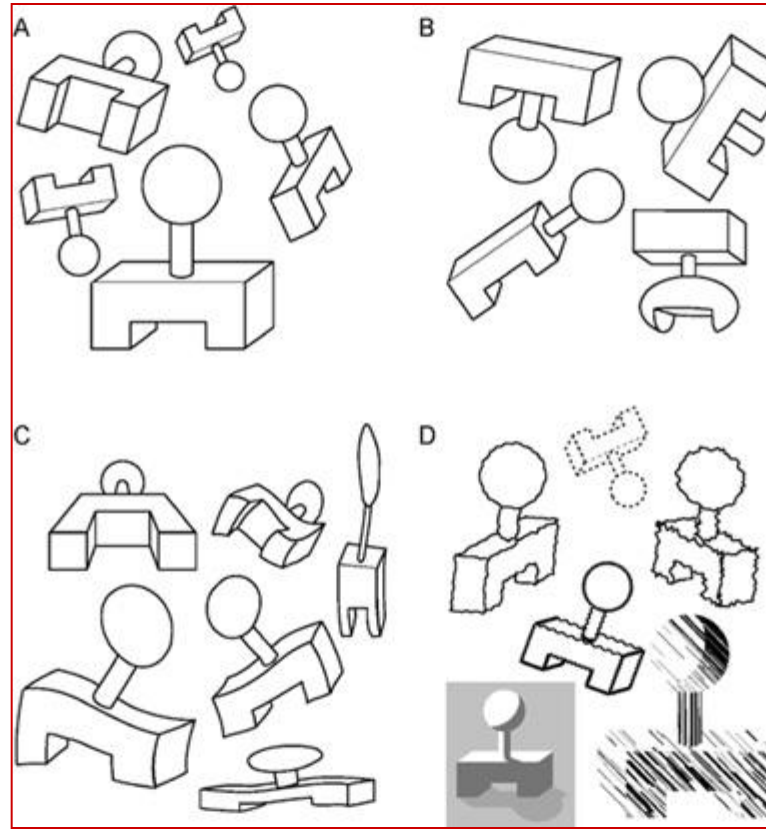
The concept of learning - example



In many cases diagnosis is like finding a dog in this picture

An example of a difficult pattern recognition problem





Invariance

Patterns and pattern classes

A **pattern** is a set of features that form a qualitative or structural description of an analysed object; in a mathematical sense a pattern is defined as a vector of features $\boldsymbol{x}=[x_1, x_2, \ldots, x_N]$.

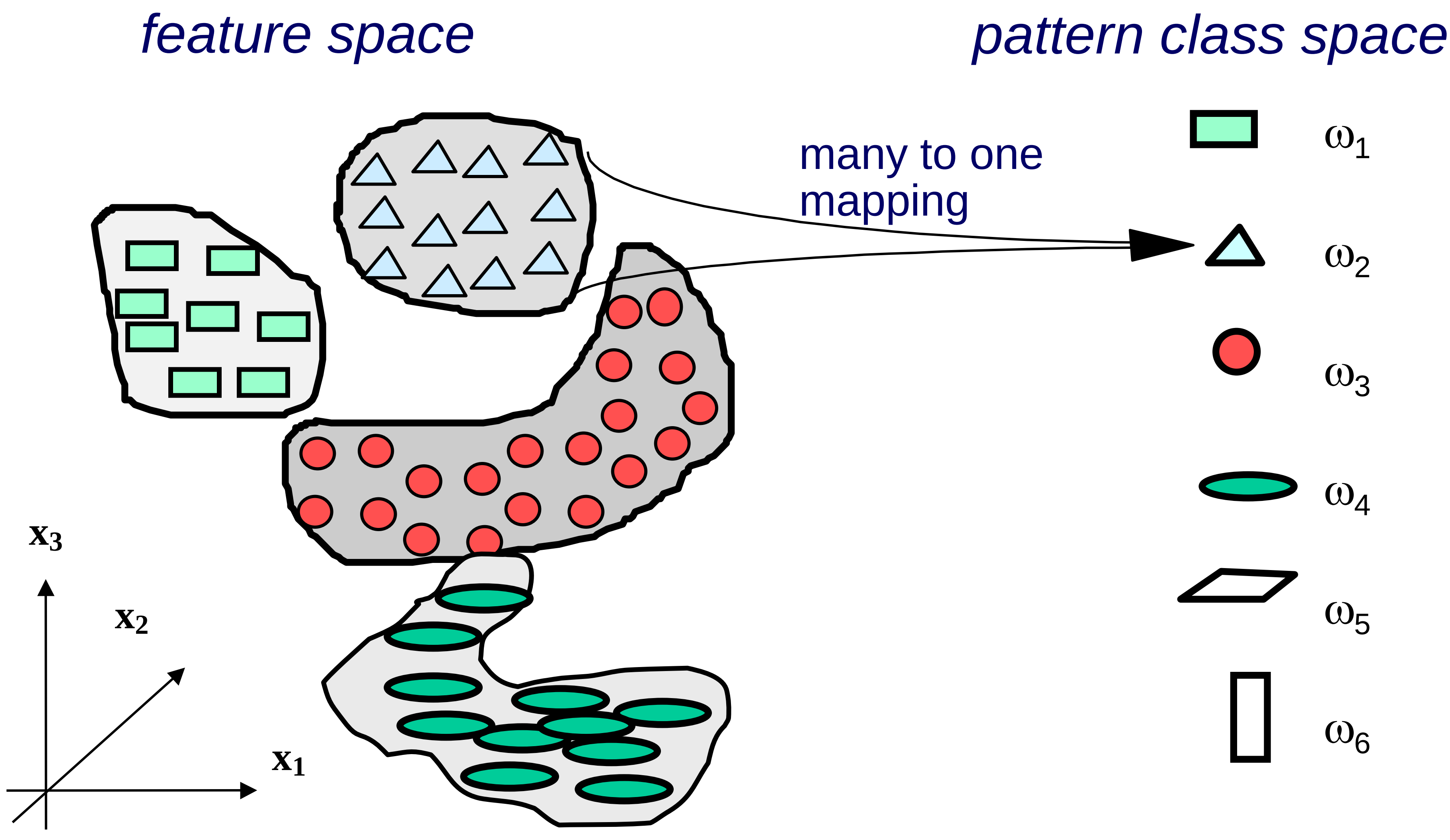
A **pattern class** is a group of patterns that have similar feature vectors (according to some similarity measure). Pattern classes are denoted $\omega_1, \omega_1, \ldots \omega_M$, where M is the number of classes.

Pattern recognition (alternatively termed pattern classification) is the task of assigning patterns to their respective classes. It is equivalent to establishing a mapping:

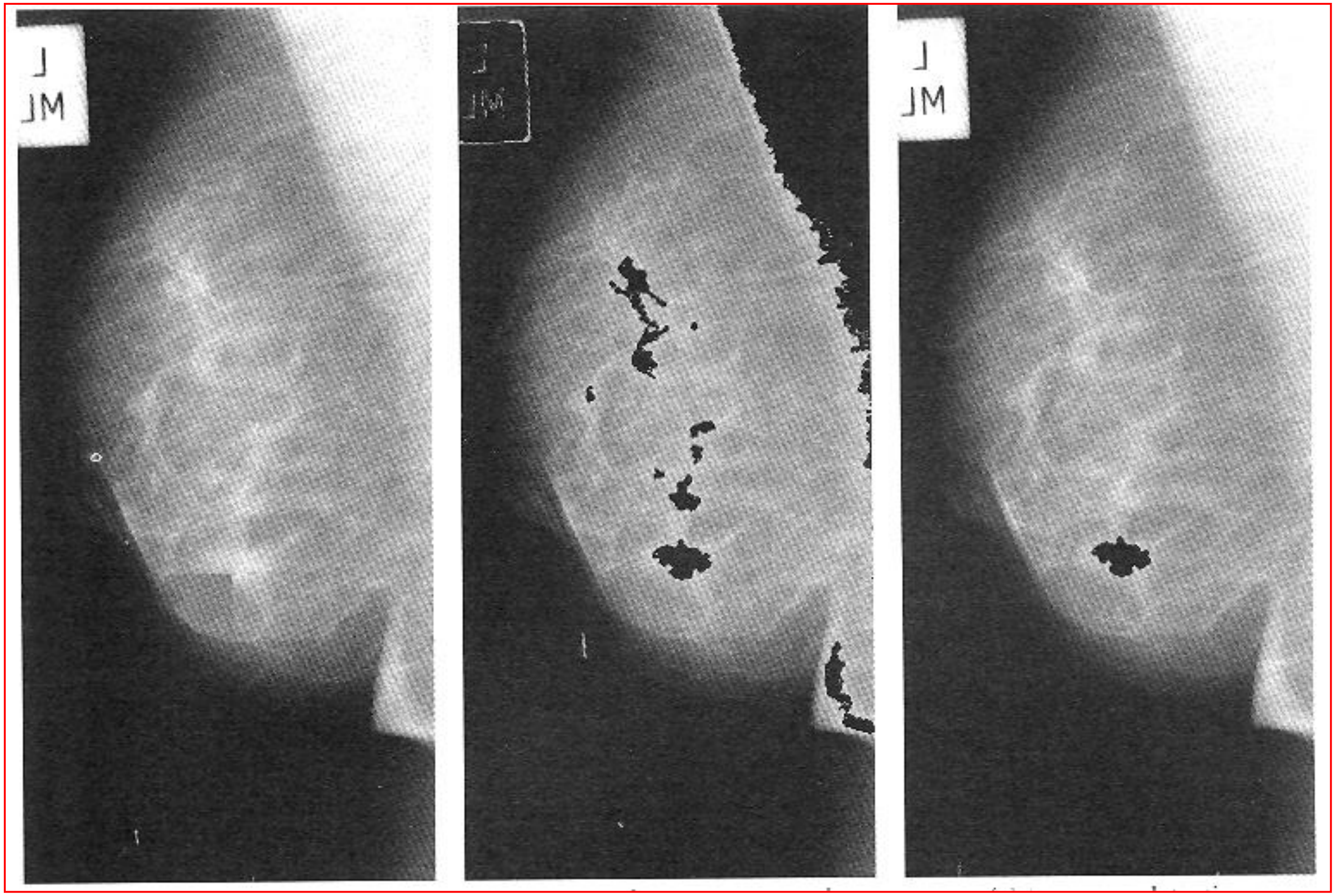
$$\boxed{\mathbf{x} \longrightarrow \omega}$$

from the **feature space** X to the **pattern class space** Ω.

Pattern classification definition



Analysis of biomedical images



Selection of feature properties

- **discrimination** – features should assume considerably different values for patterns from different classes, e.g., diameter of fruits is a good feature for classification between grapefruits and cherries,
- **robustness** – features should assume similar values for all patterns belonging to the same class, e.g., colour is a poor feature for apples,
- **independence** – features used in a classification system should be uncorrelated, e.g., weight and size of a fruit are strongly correlated features,
- **small number of features –** complexity of data classification system grows substantially with the number of classified features, e.g., features that are strongly correlated should be eliminated.

Selection of feature properties

Correlation coefficient between features *X* and *Y*:

$$\boxed{\gamma_{xy} = \frac{\frac{1}{P}\sum_{i=1}^{P}\left(x_i - \mu_x\right)\left(y_i - \mu_y\right)}{\sigma_x \sigma_y}}$$

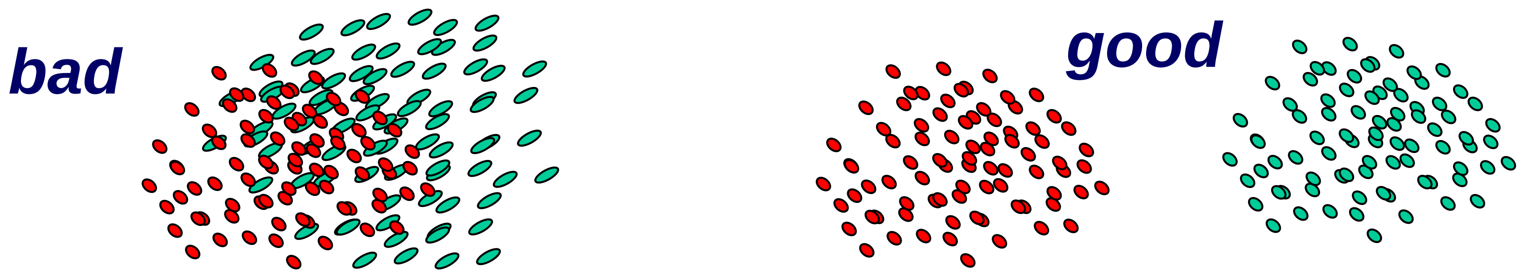
where: *P* – is the number of patterns, μ_x, μ_y are the average values for features *X* and *Y* correspondingly, and σ_x, σ_y are their standard deviations. The correlation coefficient assumes values in the range [-1, 1].

Selection of feature properties

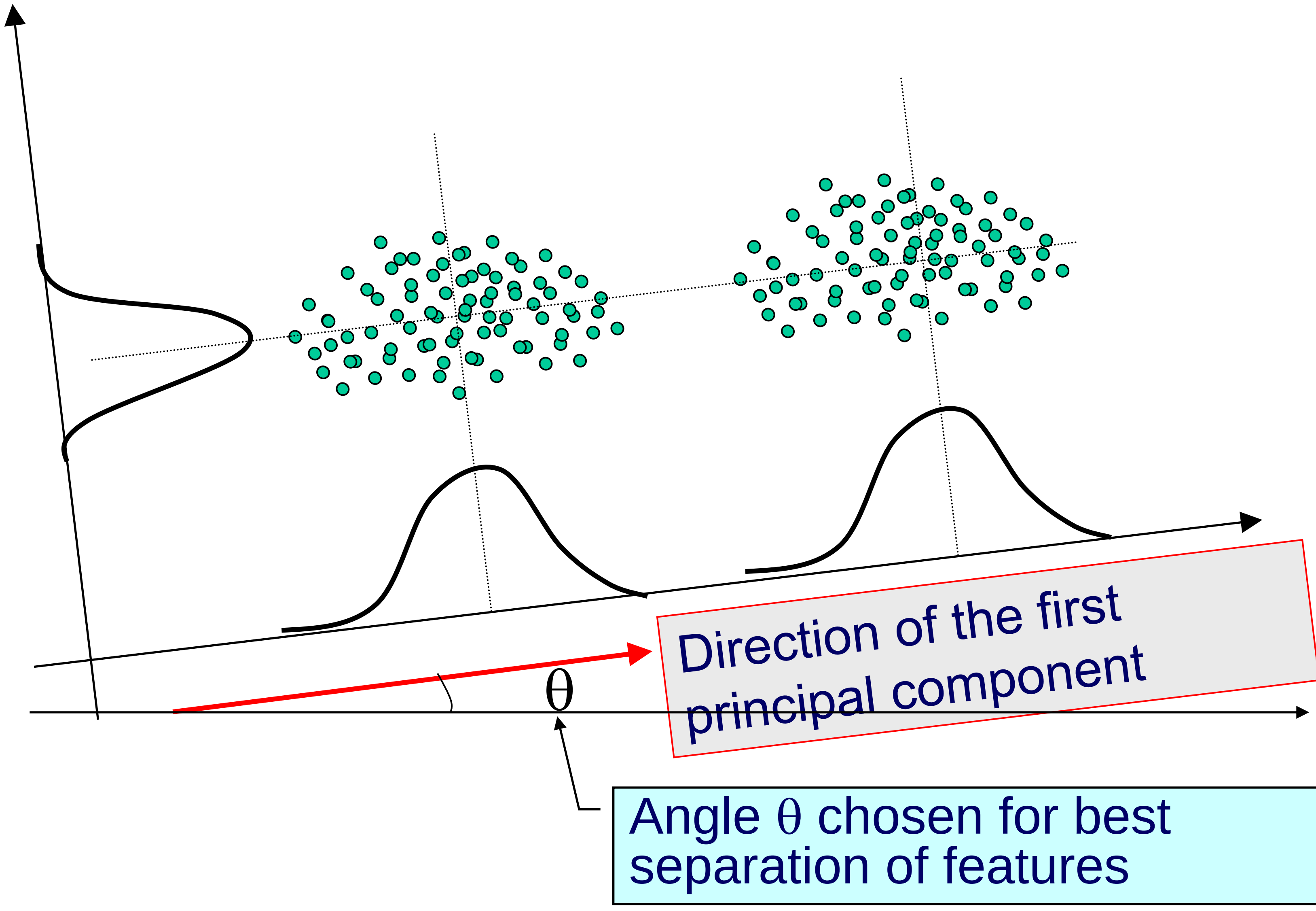
A measure of **separation** for feature *X* between classes *j* and *k*:

$$\hat{D}_{xjk} = \frac{\left|\mu_{xj} - \mu_{xk}\right|}{\sqrt{\sigma^2_{xj} + \sigma^2_{xk}}}$$

Large value for this measure means feature X yields good separation between classes.



Reduction of the number of features



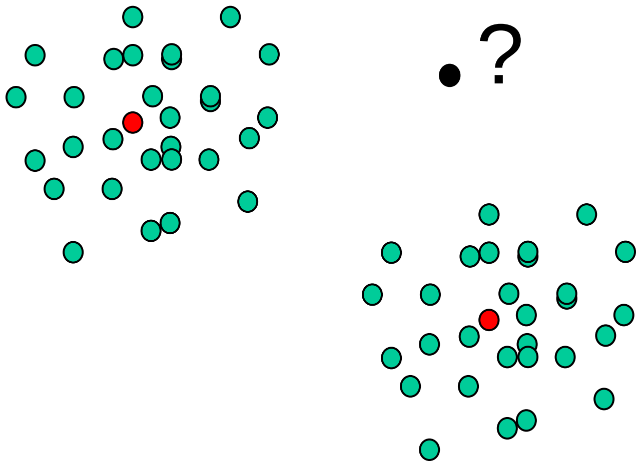
Minimum distance classifier

Assume that each pattern class is represented by a mean vector (also called a class *prototype*):

$$\boldsymbol{m}_i = \frac{1}{P_j} \sum_{x \in \omega_j} \boldsymbol{x}, \qquad j = 1,2,\ldots,M$$

where N_j is the number of pattern vectors from class ω_j.

Possible way to determine the class membership of an unknown pattern vector ***x*** is to assign it to the class of its closest prototype vector.



Minimum distance classifier

If the Euclidean distance is used the distance measure is of the form:

$$\boxed{D_j(\boldsymbol{x}) = \left\| \boldsymbol{x} - \boldsymbol{m}_j \right\| \qquad\qquad j = 1, 2, \ldots, M}$$

and $\left\| \boldsymbol{x} \right\| = (\boldsymbol{x}^T \boldsymbol{x})^{1/2}$.

Feature vector ***x*** is assigned to class ω_j if $D_j(\boldsymbol{x})$ is the smallest distance.

Minimum distance classifier

The following distance function can be constructed

$$\boxed{d_j(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{m}_j - \frac{1}{2}\boldsymbol{m}_j^T \boldsymbol{m}_j \qquad j = 1,2,\ldots,M}$$

and assigning ***x*** to class ω_j if $d_j(\boldsymbol{x})$ gives the largest value.

Minimum distance classifier

The decision boundary between classes ω_i and ω_j for a minimum distance classifier is:

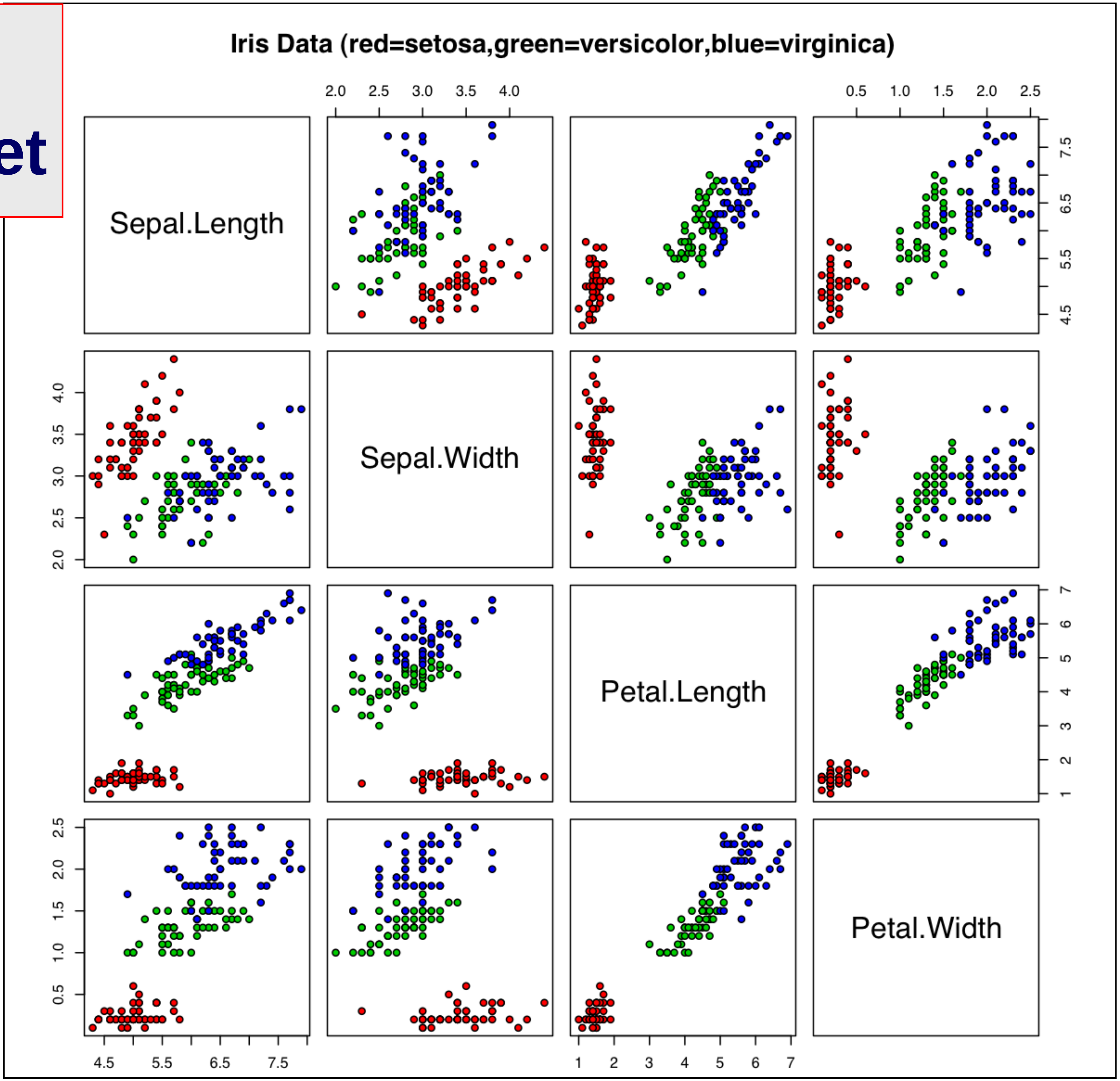
$$\boxed{\begin{aligned} & d_{ij}(\boldsymbol{x}) = d_i(\boldsymbol{x}) - d_j(\boldsymbol{x}) = \\ & = \boldsymbol{x}^T(\boldsymbol{m}_i - \boldsymbol{m}_j) - \frac{1}{2}(\boldsymbol{m}_i - \boldsymbol{m}_j)^T(\boldsymbol{m}_i - \boldsymbol{m}_j) = 0 \end{aligned}}$$

The surface defined by this equation is the perpendicular bisector to the line joining $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$. For $N=2$ the bisector is a line, for $N=3$ it is a plane, and for $N>3$ it is called a *hyperplane*.

Fisher's Iris data set

Setosa *Virginica* *Versicolor*

Fisher's Iris data set



Fisher's Iris data set

f1	f2	f3	f4	
4.9	3.0	1.4	0.2	*setosa*
4.7	3.2	1.3	0.2	*setosa*
…	…	…	…	…
6.2	2.2	4.5	1.5	*versicolor*
5.9	3.2	4.8	1.8	*versicolor*
…	…	…	…	…
6.4	2.7	5.3	1.9	*virginica*
5.7	2.5	5.0	2.0	*virginica*

150 sets; 50 for each Iris species

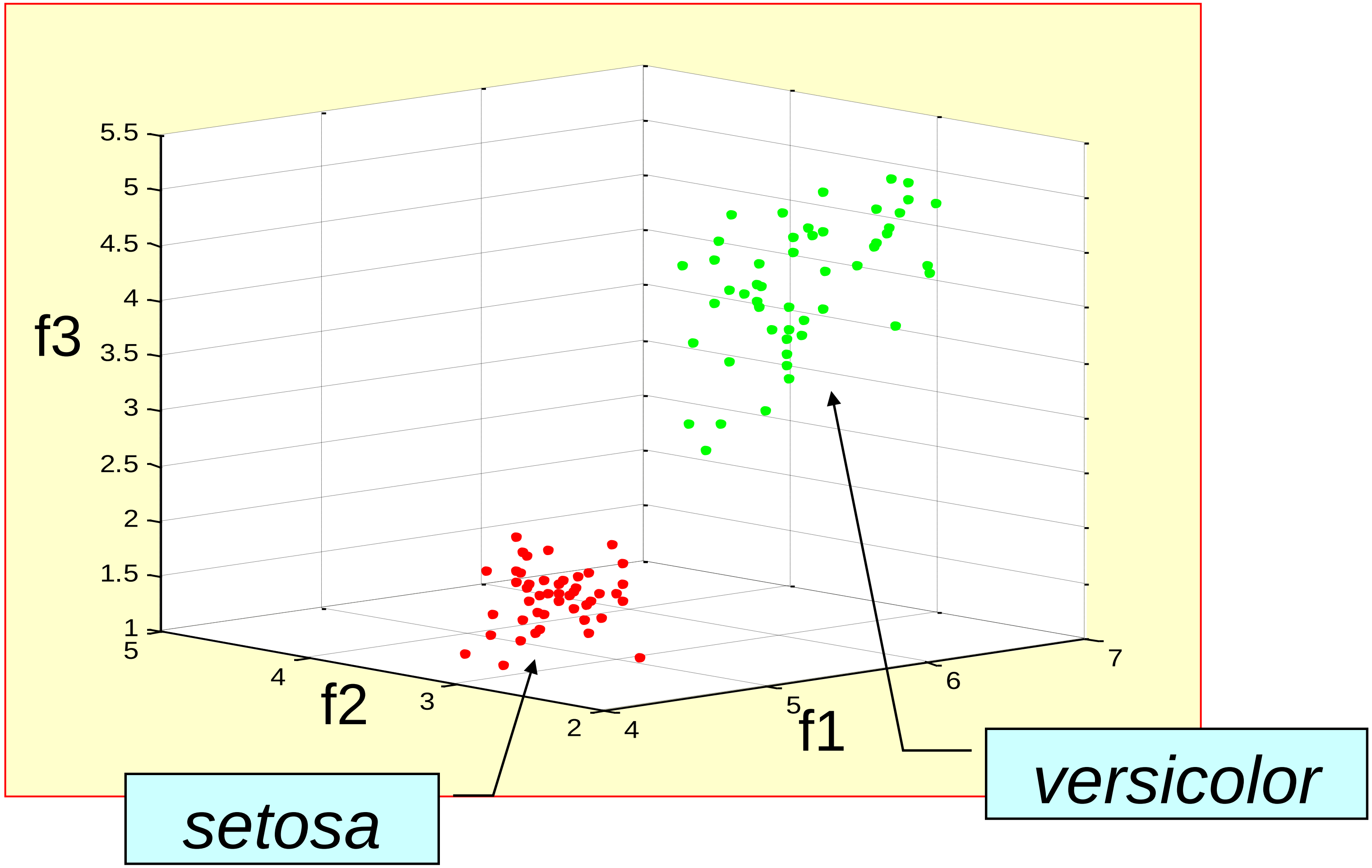
Petal width

Petal length

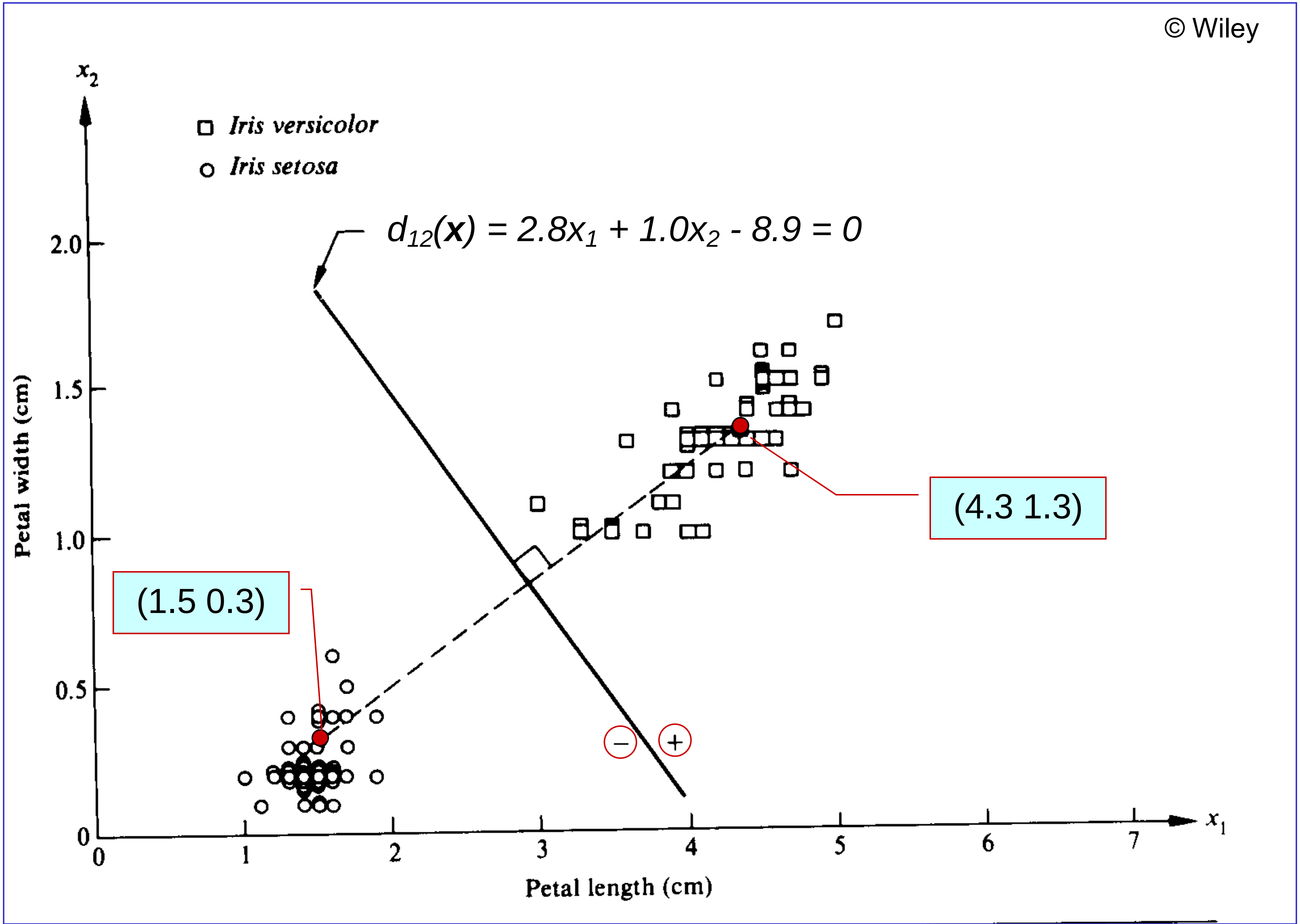
Sepal width

Sepal length

Fisher's Iris data set



Minimum distance classifier



Minimum distance classifier

Example 1: Consider two classes denoted by ω_1 and ω_2, that have mean vectors $\boldsymbol{m}_1 = (4.3,\ 1.3)^T$ and $\boldsymbol{m}_2 = (1.5,\ 0.3)^T$. The decision functions for each of the classes are:

$$d_1(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{m}_1 - 0.5\ \boldsymbol{m}_1^T\boldsymbol{m}_1 = 4.3x_1 + 1.3x_2 - 10.1$$

$$d_2(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{m}_2 - 0.5\ \boldsymbol{m}_2^T\boldsymbol{m}_2 = 1.5x_1 + 0.3x_2 - 1.17$$

Equation for the decision boundary :

$$d_{12}(\boldsymbol{x}) = d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}) = 2.8x_1 + 1.0x_2 - 8.9 = 0$$

Class membership of a new feature vector is defined on the basis of the sign of $d_{12}(\boldsymbol{x})$.

Fisher's Iris data set

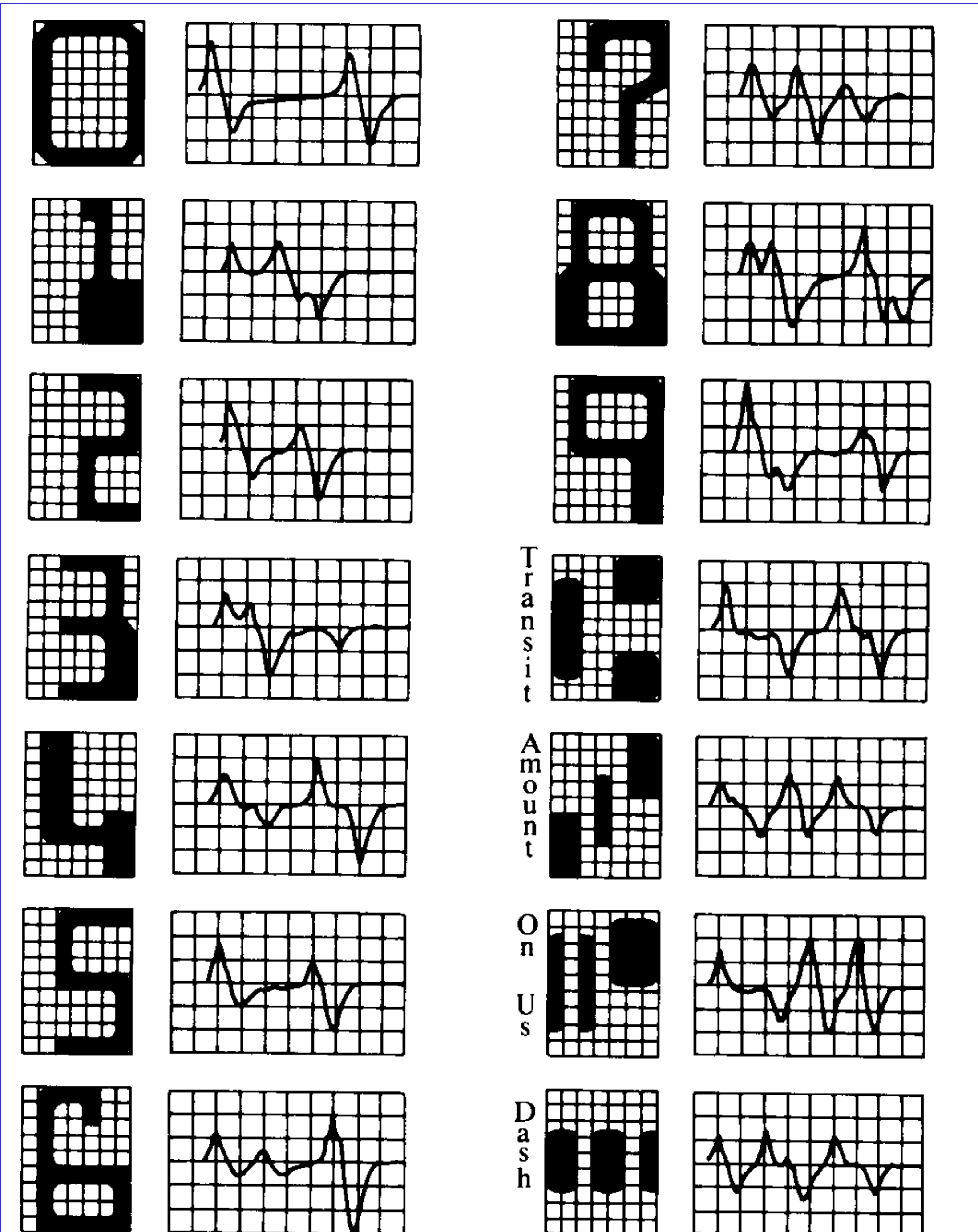
Exercise: Recall the class separability measure for feature x between classes j and k:

$$\hat{D}_{xjk} = \frac{\left|\mu_{xj} - \mu_{xk}\right|}{\sqrt{\sigma_{xj}^2 + \sigma_{xk}^2}}$$

Compute this measure for each of the features and all two-class combination of Irises. Indicate the best and the worst separable features. Compare it with Iris feature plots in earlier slide.

Fisher's Iris data set

	Sepal length	*Sepal width*	*Petal length*	*Petal width*
Setosa - *Virginica*				
Setosa - *Versicolor*				
Versicolor - *Virginica*				



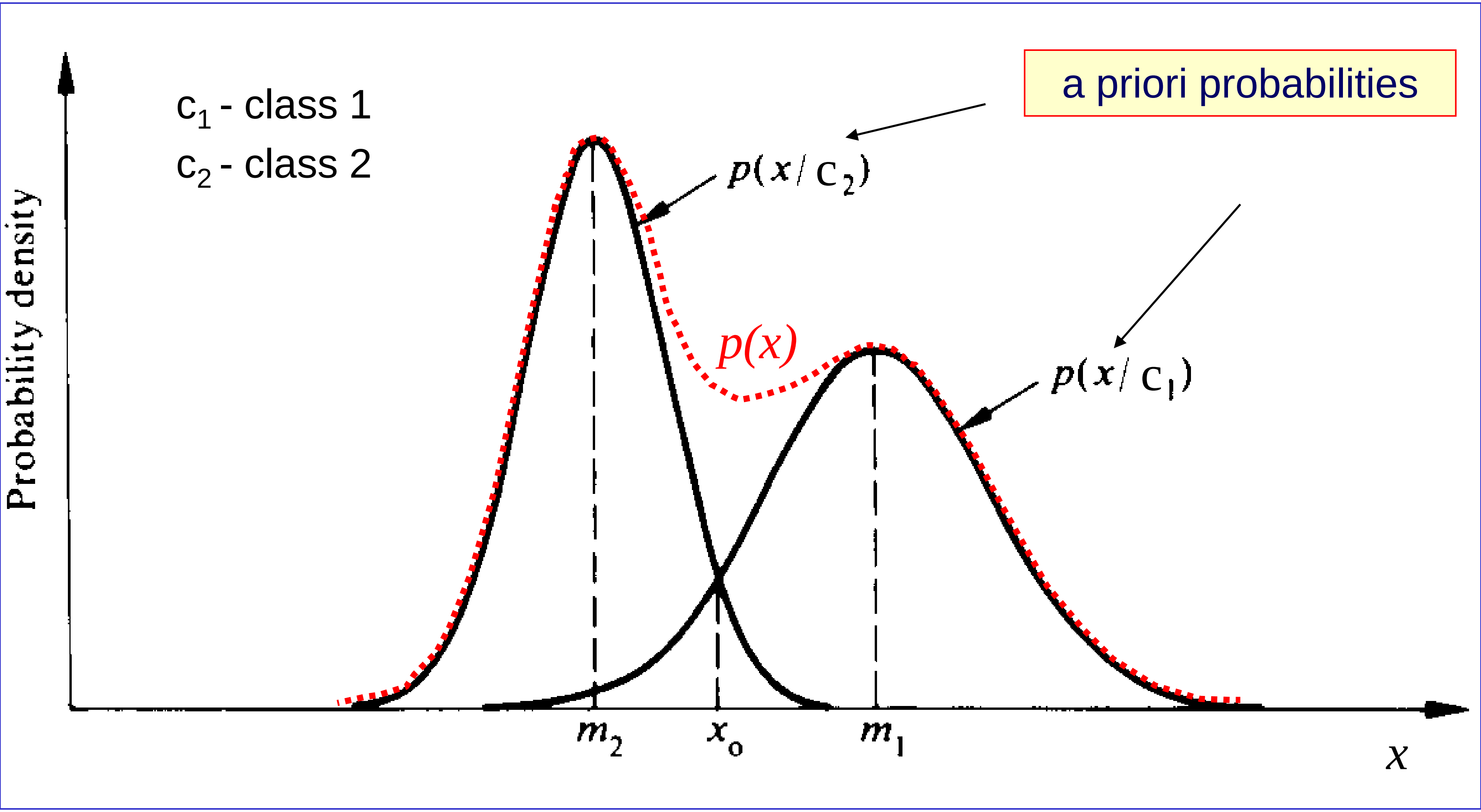
Example 2:

Classification of American Bankers Association font character set.

The design of the font ensures that the waveform corresponding to each character is distinct from that of all others.

There are N=14 prototype vectors in 10-D feature space.

The Bayes classifier



The Bayes classifier

From the probability theory the following holds:

$$\boxed{p(a/b) = \frac{p(a)\,p(b/a)}{p(b)}}$$

hence:

a priori probability

$$\boxed{p\left(c_i / \boldsymbol{x}\right) = \frac{P\left(c_i\right)p\left(\boldsymbol{x} / c_i\right)}{p\left(\boldsymbol{x}\right)}}$$

Bayes rule

a posteriori probability

Maximum likelihood estimator

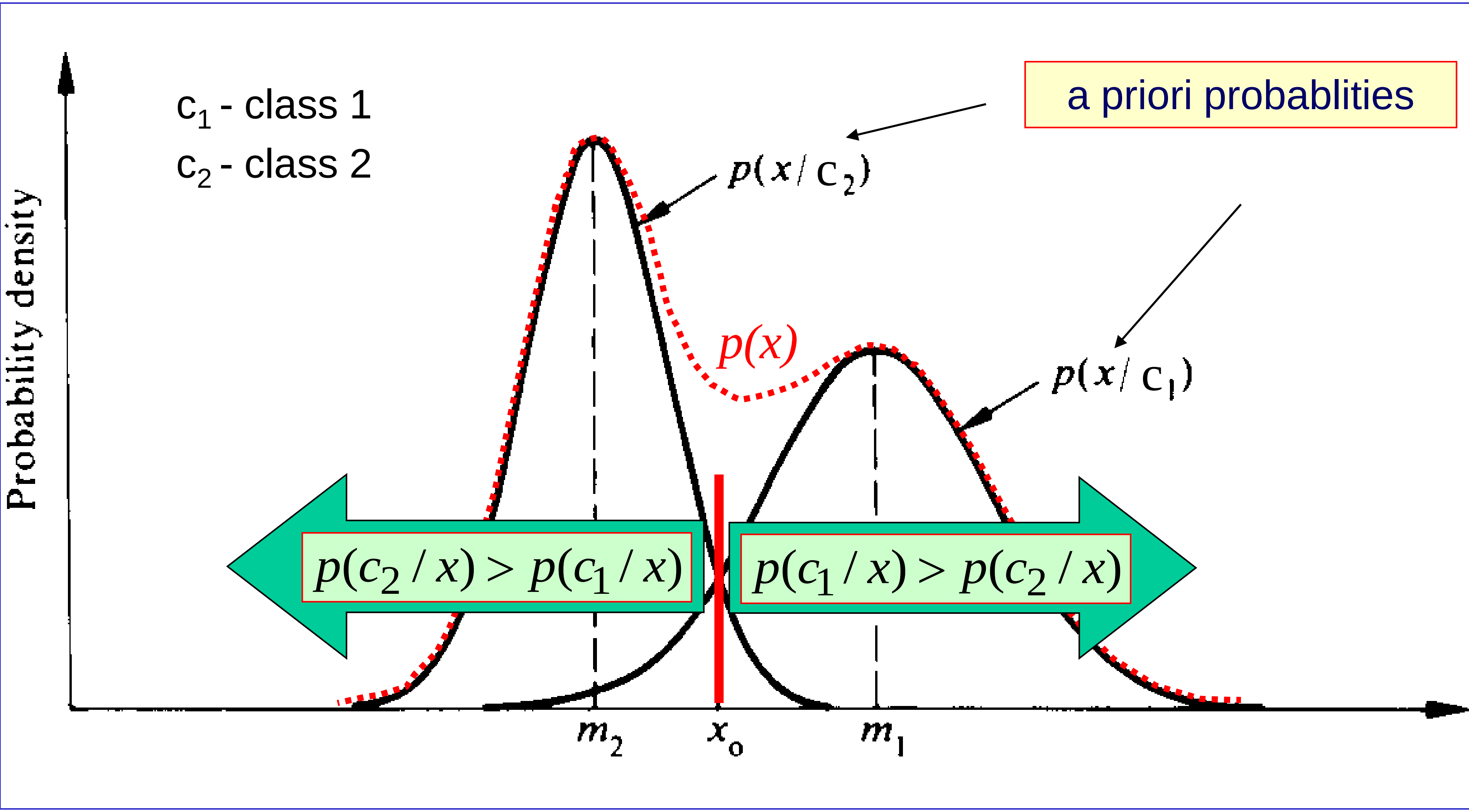
A new pattern is assigned to i-th class for which:

$$\boxed{L_i = \arg\max_i \left\{ p\left(c_j / \boldsymbol{x}\right) = \frac{P\left(c_i\right) p\left(\boldsymbol{x} / c_i\right)}{p\left(\boldsymbol{x}\right)} \right\}, \; i = 1,2,\ldots,N}$$

i.e.

$$\boxed{L_i > L_j, \;\; j \neq i, i = 1,2,\ldots,N}$$

The Bayes classifier



The Bayes classifier

A priori probability for 1D Gaussian distribution:

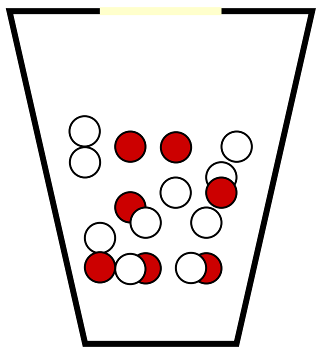
$$\boxed{p(x/c_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(x-m_i)^2}{2\sigma_i^2}\right] \quad i=1,2}$$

A posteriori probability:

$$\boxed{p(c_i/x) = p(x/c_i)P(c_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(x-m_i)^2}{2\sigma_i^2}\right]P(c_i) \qquad i=1,2}$$

The Bayes classifier - example

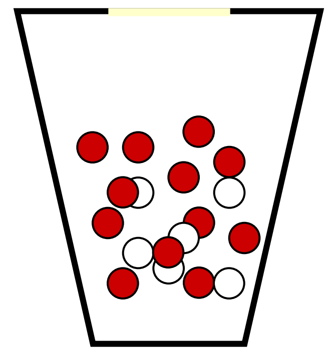
Pot A



20 – red

20 – white

Pot B



30 – red

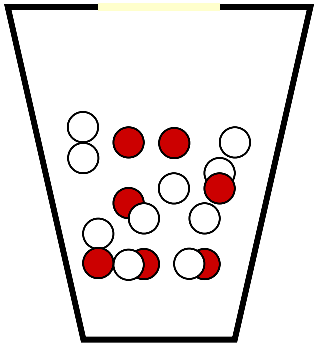
10 – white

We pick a pot randomly and then the ball randomly.

Question: From which pot the ball was picked?

The Bayes classifier - example

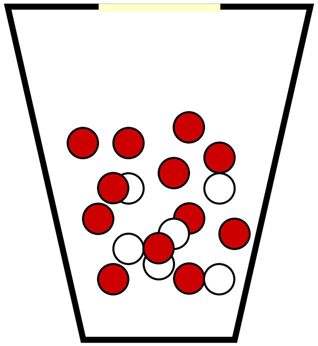
Pot A



20 – red

20 – white

Pot B



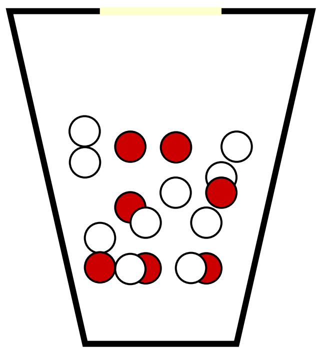
30 – red

10 – white

Before we see the colour of the ball we picked (a priori knowledge) we assume the probability of choosing each of the pots is equal, i.e. 0.5.

The Bayes classifier - example

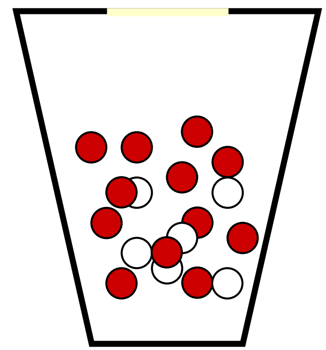
Pot A



20 – red

20 – white

Pot B



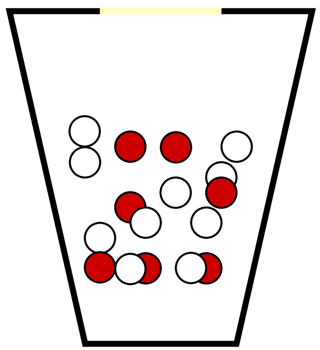
30 – red

10 – white

Suppose we picked a red ball. Can we verify our first hypothesis having this a posteriori knowledge?
Yes, the answer comes from the Bayes theorem.

The Bayes classifier -example

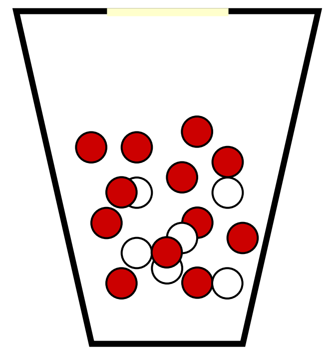
Pot A



20 – red

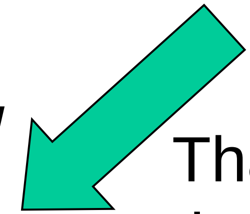
20 – white

Pot B



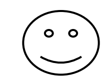
30 – red

10 – white



That was the pot!

probably



$$p(A/\textbf{red}) = \frac{P(A)p(\textbf{red}/A)}{P(A)p(\textbf{red}/A)+P(B)p(\textbf{red}/B)} = \frac{0.5\times 0.5}{0.5\times 0.5+0.5\times 0.75} = 0.4$$

$$p(B/\textbf{red}) = \frac{P(B)p(\textbf{red}/B)}{P(A)p(\textbf{red}/A)+P(B)p(\textbf{red}/B)} = \frac{0.5\times 0.75}{0.5\times 0.5+0.5\times 0.75} = 0.6$$

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

„Signal processing”

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego w projekcie pt.

„Innowacyjna dydaktyka bez ograniczeń - zintegrowany rozwój Politechniki Łódzkiej - zarządzanie Uczelnią, nowoczesna oferta edukacyjna i wzmacniania zdolności do zatrudniania osób niepełnosprawnych”