

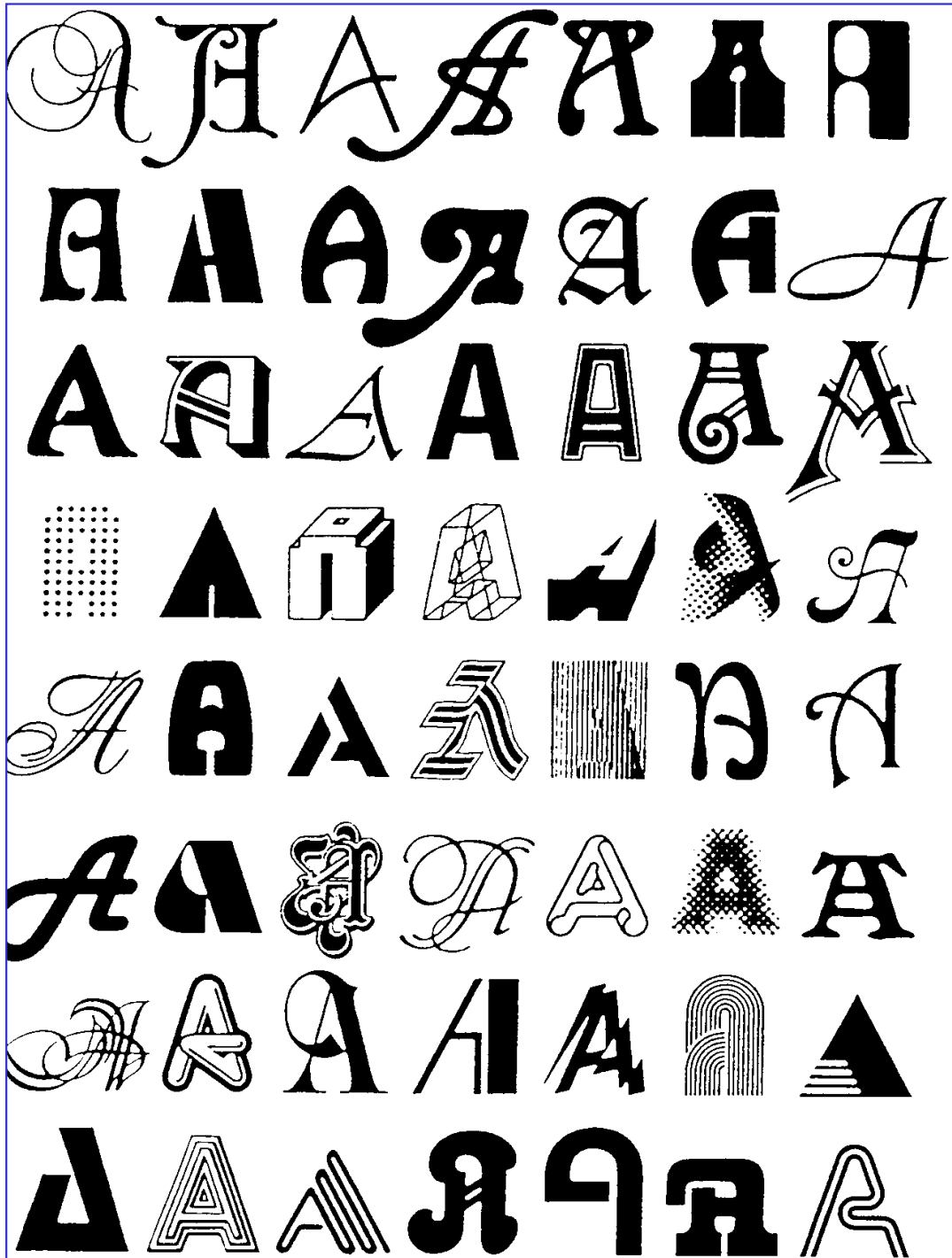
Klasyfikacja danych – wprowadzenie

Przykład trudnego problemu rozpoznawania obrazu

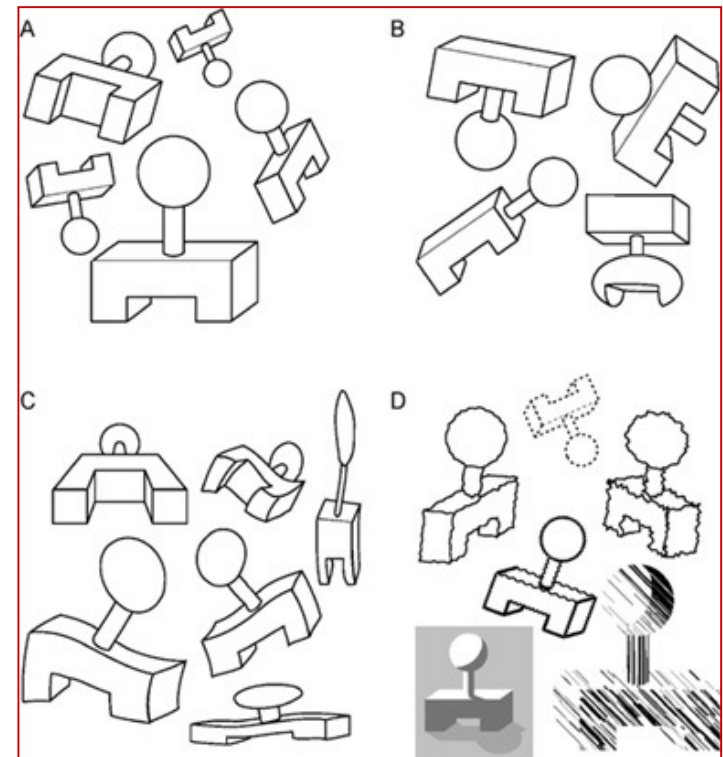


Wyłanianie

Paweł Strumiłło, Instytut Elektroniki Politechniki Łódzkiej

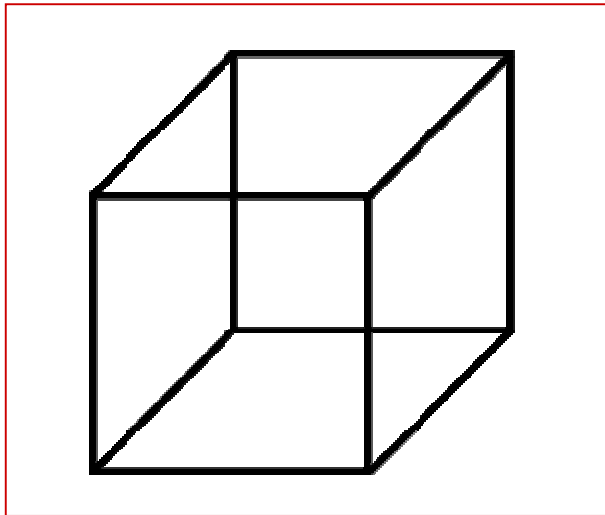
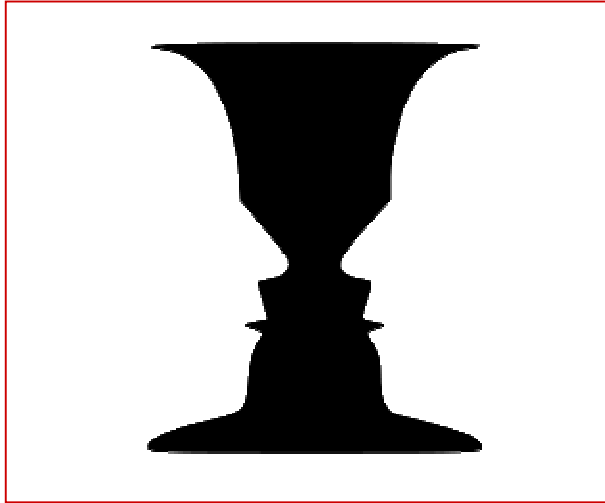


Przykład trudnego zadania rozpoznawania obrazu (wymagającego znamion zachowania inteligentnego)

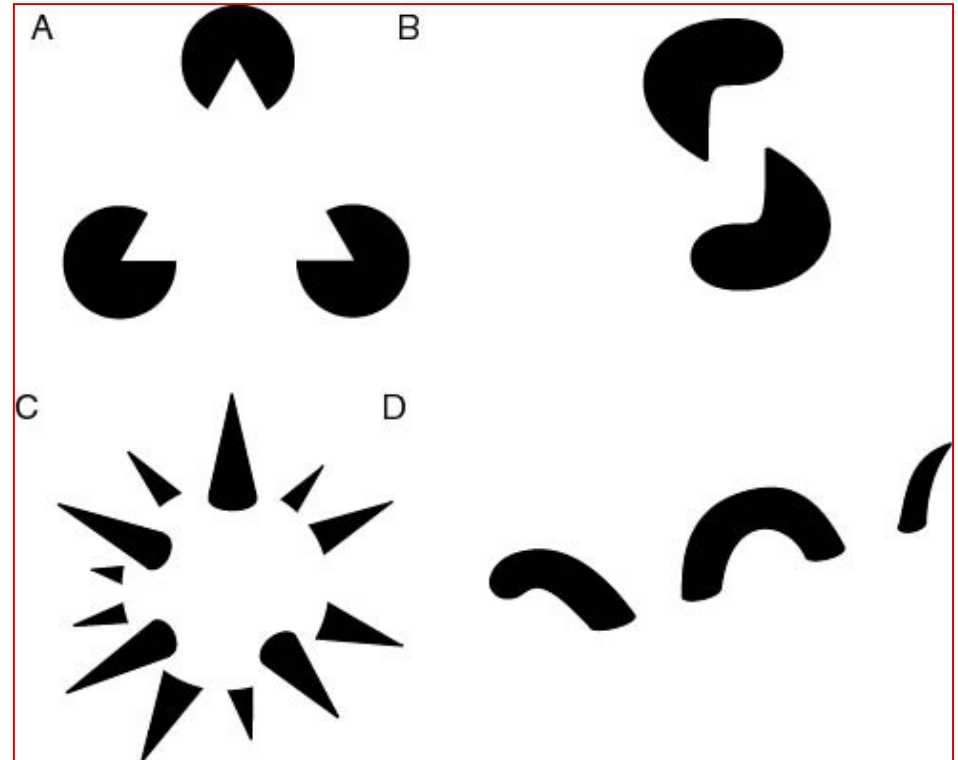


Niezmiennność

Teoria Gestalt (niem. *postać*)



Wieloznaczność



Reifikacja
(iluzoryczne obiekty)

Wzorce i klasy wzorców

Wzorzec to zbiór cech, który tworzy ilościowy i jakościowy opis obiektu; ściślej, wzorzec to wektor cech $\mathbf{x}=[x_1, x_2, \dots, x_N]$.

Klasa wzorców to zbiór wzorców charakteryzujących się podobnymi wektorami cech. Klasy wzorców oznaczmy $\omega_1, \omega_2, \dots, \omega_M$ gdzie indeks M jest numerem klasy.

Rozpoznawanie wzorców (nazywane też klasyfikacją) jest zadaniem polegającym na przyporządkowaniu wzorców do ich klas:

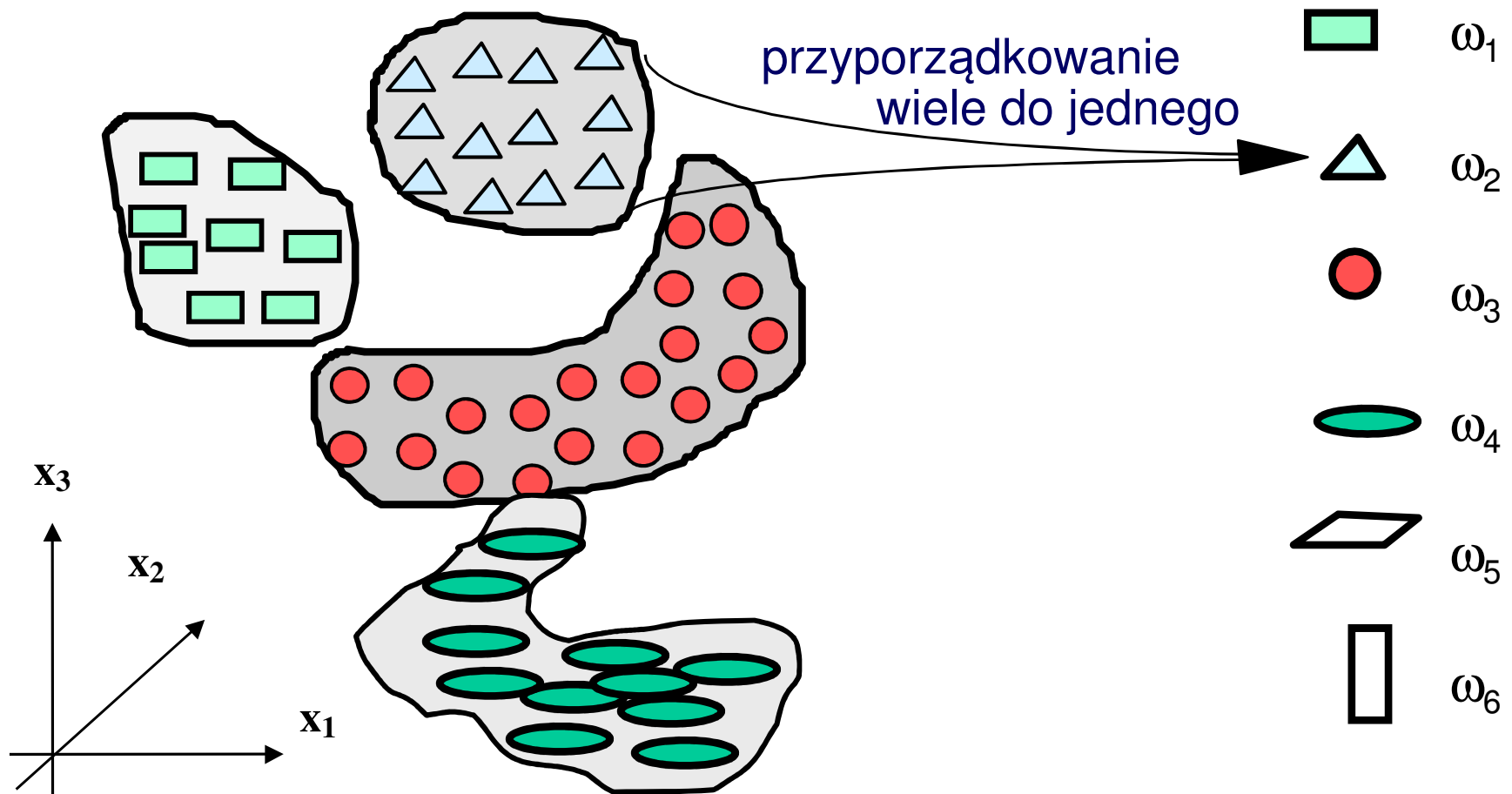
$$\mathbf{x} \rightarrow \omega$$

tj. przekształceniem przestrzeni wektorów cech X na przestrzeń klas wzorców Ω .

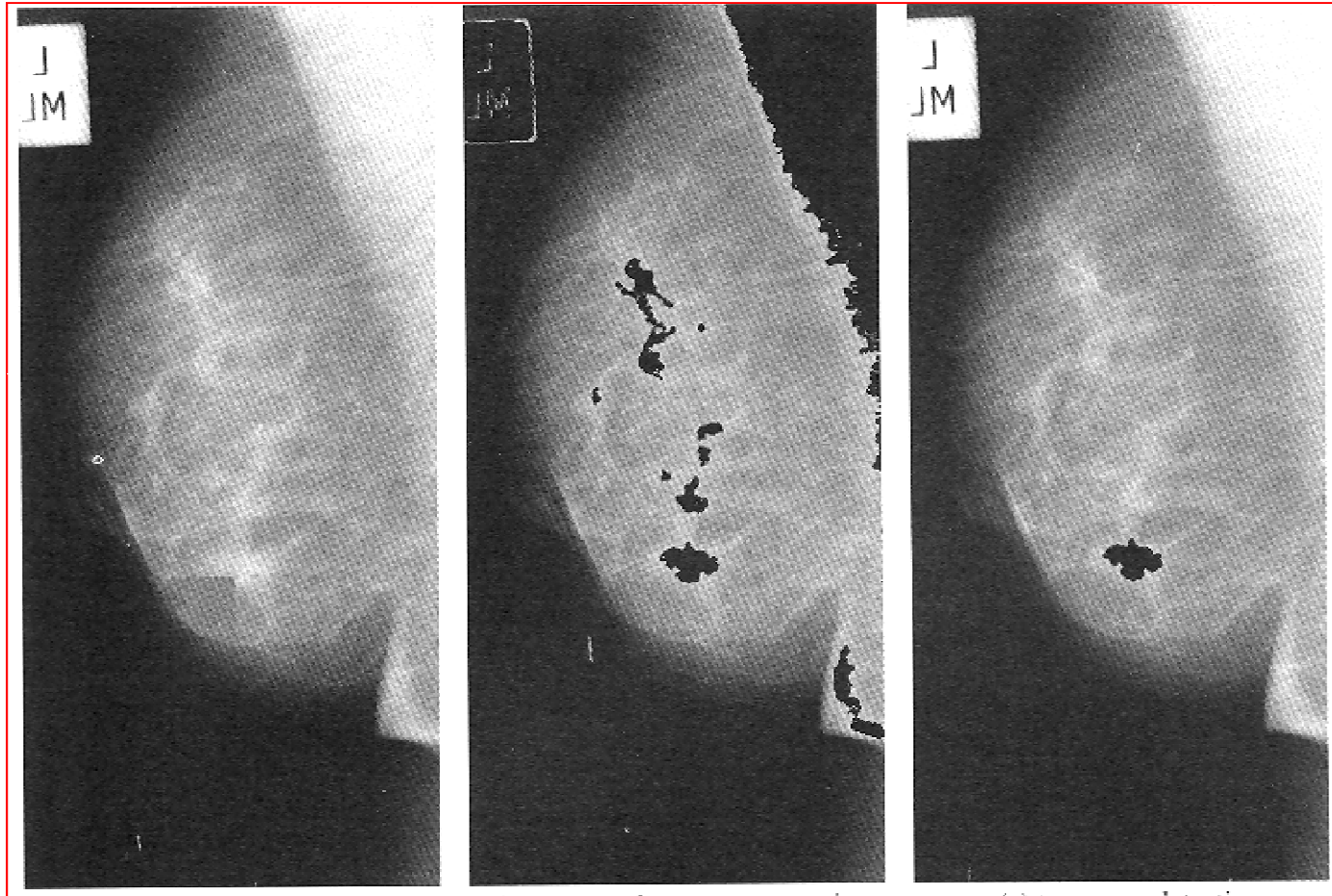
Klasyfikacja wzorców

Przestrzeń cech

Przestrzeń klas



Analysis of biomedical images



Paweł Strumiłło, Instytut Elektroniki Politechniki Łódzkiej

Selekcja cech i ich właściwości

- **dyskryminacja** - cechy powinny przyjmować znacząco różne wartości dla obiektów z różnych klas, np. średnica owocu jest dobrą cechą dla rozróżnienia wiśni i grejfrutów,
- **niezawodność** - cechy powinny przyjmować podobne wartości dla wszystkich obiektów danej klasy, np. kolor jest złą cechą dla jabłek,
- **niezależność** - cechy wykorzystywane w danym systemie klasyfikacji powinny być nieskorelowane ze sobą, np. waga i wielkość owocu są cechami silnie skorelowanymi,
- **mała liczba** - złożoność systemu klasyfikacji rośnie szybko wraz z liczbą klasyfikowanych cech, np. należy eliminować cech skorelowane.

Selekcja cech i ich właściwości

Współczynnik korelacji cech x i y :

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

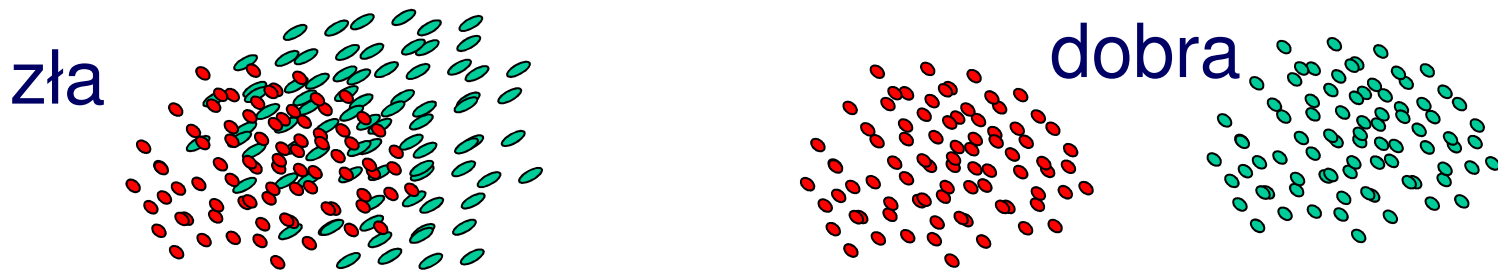
gdzie: P - liczna klasyfikowanych obiektów, μ , σ oznaczają odpowiednio wartości średnie i odchylenie standardowe danego zbioru cech. Jeżeli wsp. korelacji jest bliski 1 (-1) cechy x i y uważa się za silnie skorelowane (np. jedną z nich można odrzucić).

Selekcja cech i ich właściwości

Przykład miary separacji cechy x pomiędzy klasami j i k :

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

Duża wartość tej miary świadczy o dobrej separacji cechy x pomiędzy klasami j i k .



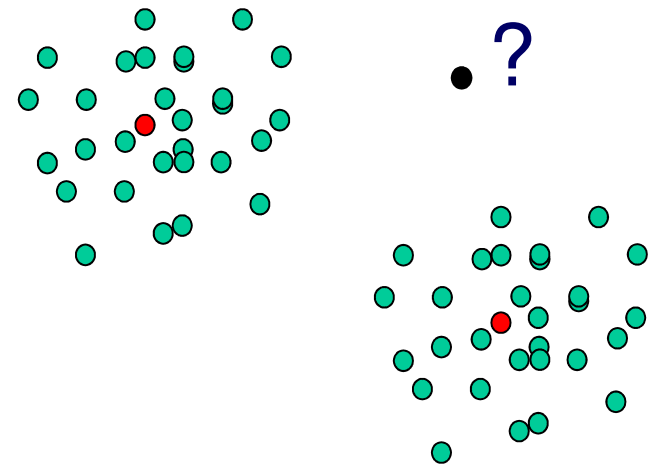
Klasyfikator minimalnoodległościowy (*ang. minimum distance classifier*)

Założmy, że każda klasa wzorów jest reprezentowana przez typowy wzorzec, zwany prototypem klasy:

$$m_i = \frac{1}{P_j} \sum_{x \in \omega_j} x_j, \quad j = 1, 2, \dots, M$$

gdzie P_j jest liczbą wzorców z klasy ω_j .

Jednym z możliwych sposobów klasyfikacji nowego wzorca x jest przydzielenie go do klasy, której prototyp znajduje się najbliżej tego wzorca.



Klasyfikator minimalnoodległościowy

Dla Euklidesowej miary odległości:

$$D_j(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_j\| \quad j = 1, 2, \dots, M$$

gdzie $\|\mathbf{z}\| = (\mathbf{z}^T \mathbf{z})^{1/2}$. Wzorzec \mathbf{x} jest przydzielony do klasy ω_j jeżeli $D(\mathbf{x}_j)$ jest odległością najmniejszą.

Klasyfikator minimalnoodległościowy

Na podstawie poprzednich zależności, funkcja decyzyjna jest postaci:

$$d_j(\mathbf{x}) = \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j \quad j = 1, 2, \dots, M$$

Wektor \mathbf{x} jest przyporządkowywany do klasy ω_j jeżeli $d_j(\mathbf{x})$ jest największe.

Klasyfikator minimalnoodległościowy

Granica decyzyjna pomiędzy klasami ω_i i ω_j dla klasyfikatora minimalnoodległościowego:

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

Obszar wyznaczony przez miejsca geometryczne spełniające to równanie dzieli na pół i jest prostopadły do linii łączącej \mathbf{m}_i i \mathbf{m}_j . Dla $N=2$ obszar ten redukuje się do linii prostej, dla $N=3$ jest płaszczyzną, oraz dla $N>3$ jest nazywany hiperpłaszczyzną.

Zbiór testowy Fisher'a

Setosa



Virginica



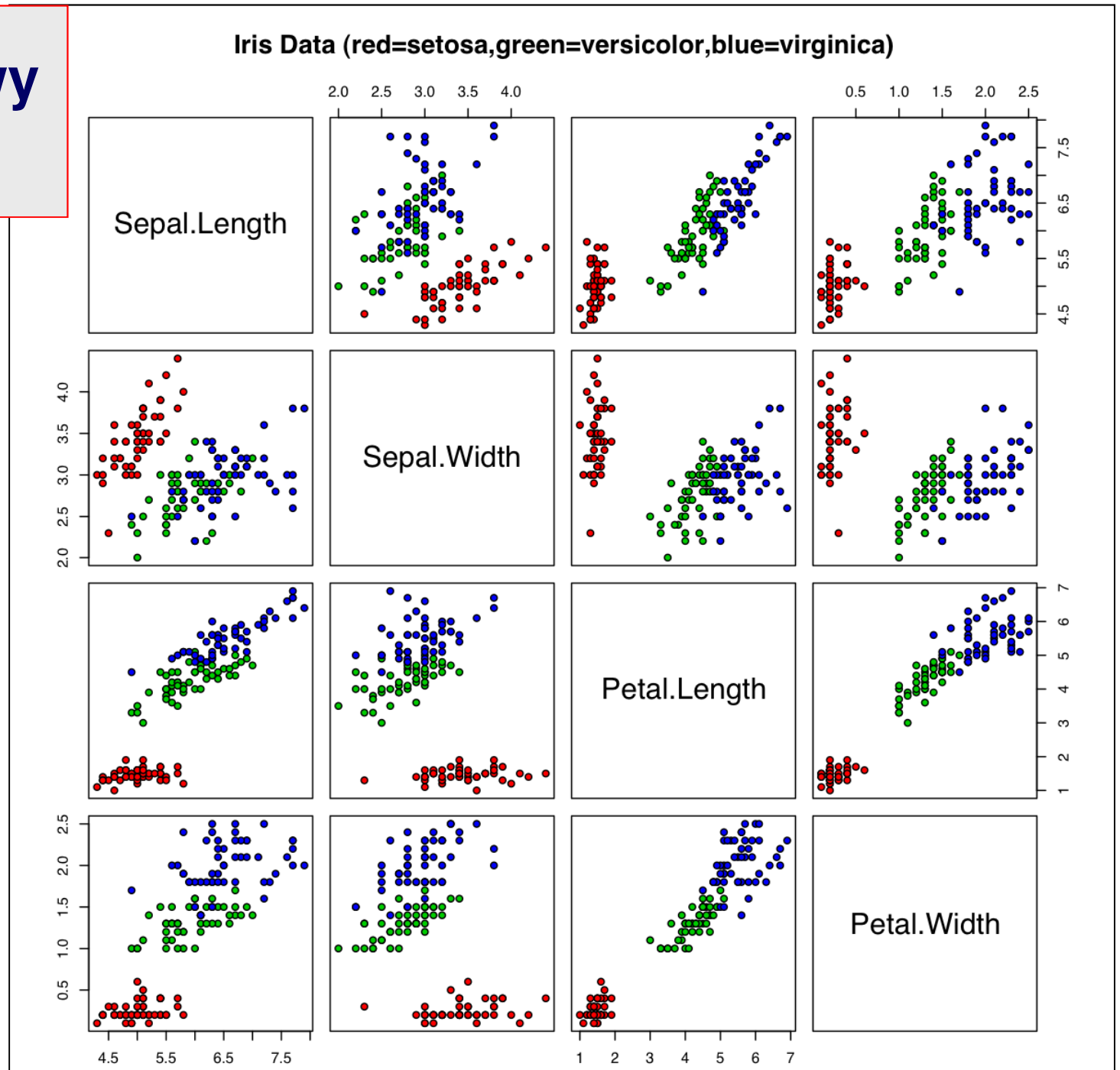
Versicolor



sepal

petal

Ziór testowy Fisher'a



Zbiór testowy Fisher'a

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

150 wektorów danych po dla każdego gatunku Irysa

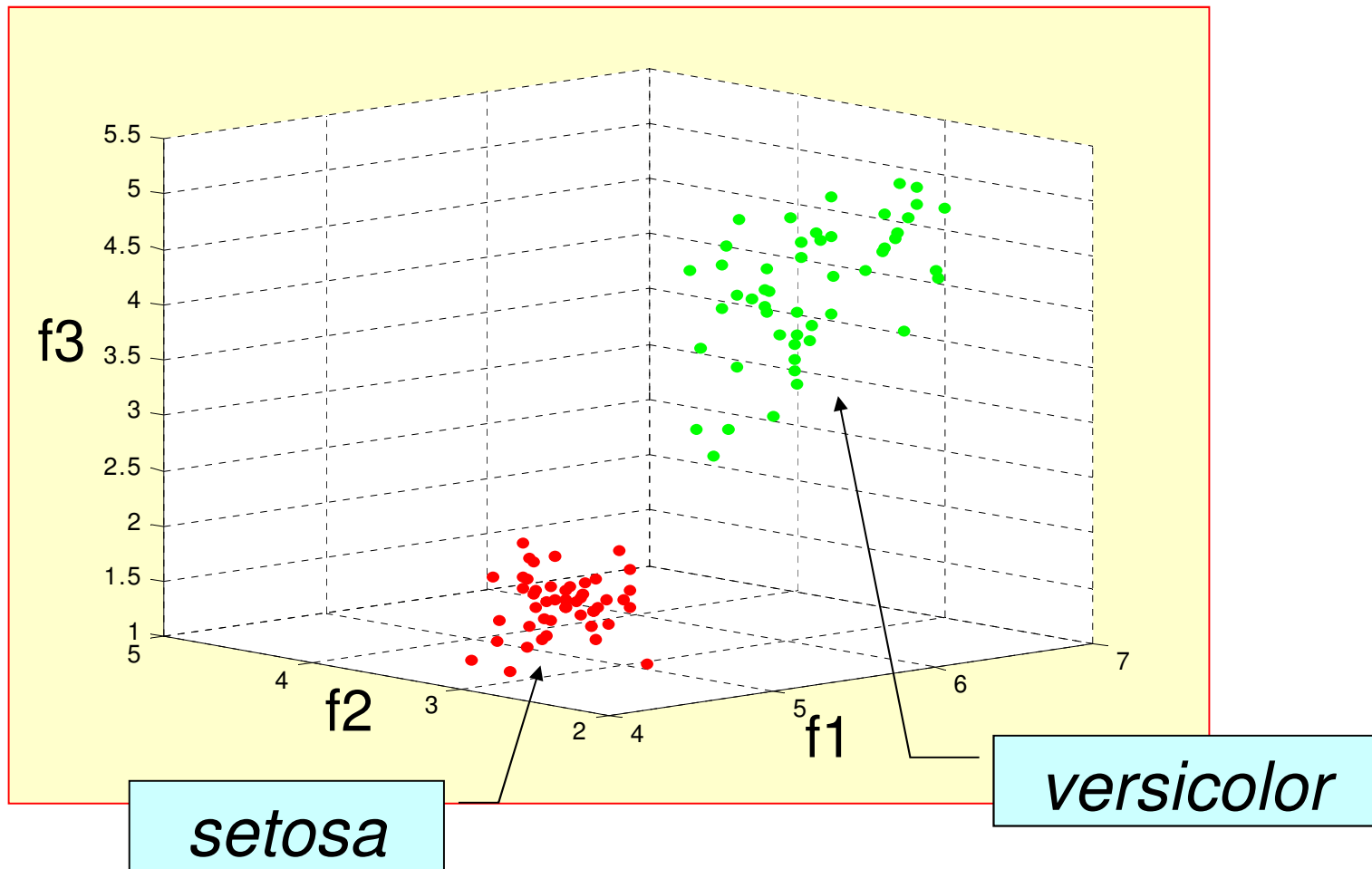
Petal width

Petal length

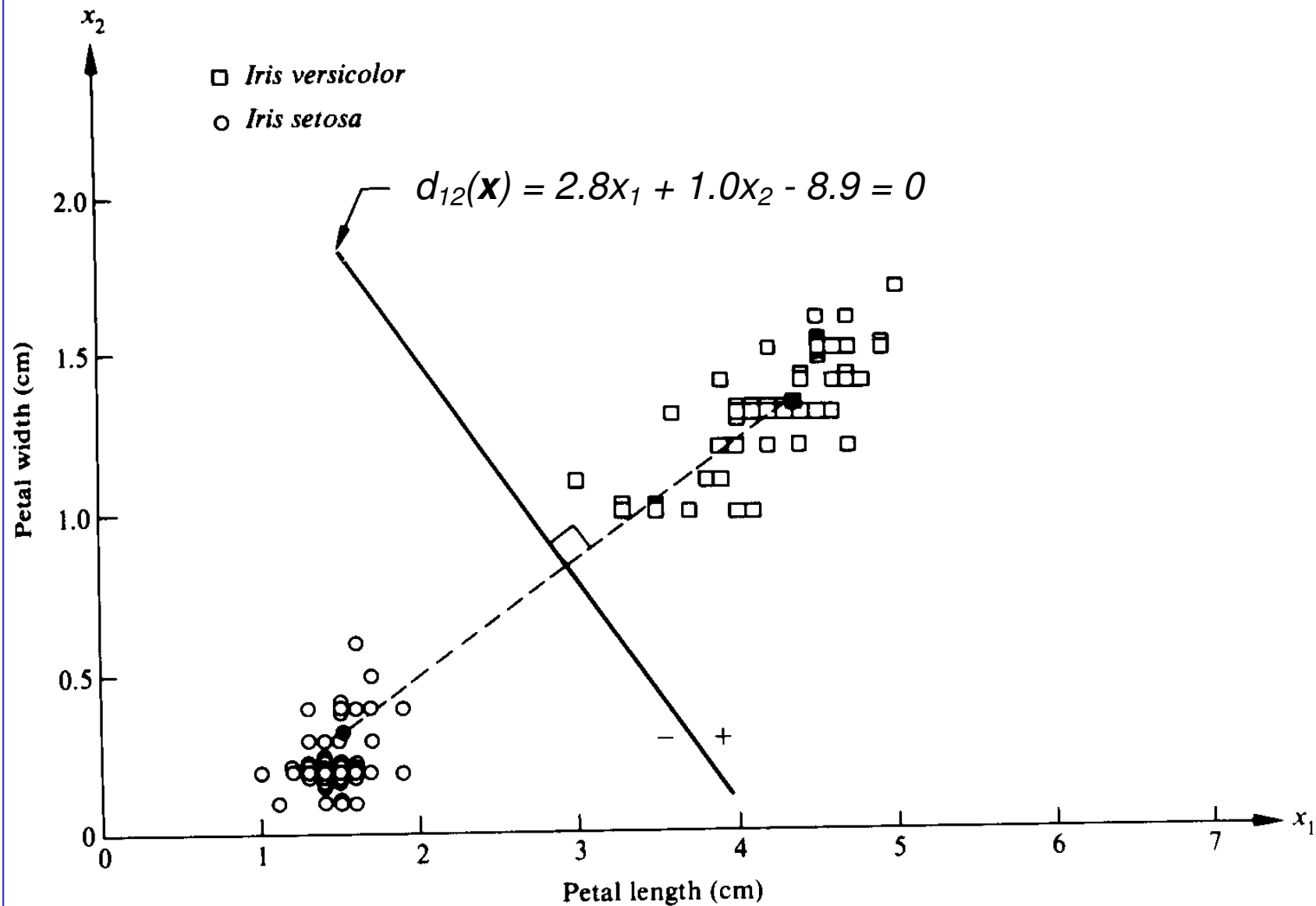
Sepal width

Sepal length

Zbiór testowy Fisher'a



Klasyfikator minimalnoodległościowy



Klasyfikator minimalnoodległościowy

Przykład 1: Rozpatrzmy dwie klasy ω_1 i ω_2 , posiadające wektory średnie cech $\mathbf{m}_1 = (4.3, 1.3)^T$ i $\mathbf{m}_2 = (1.5, 0.3)^T$. Funkcje decyzyjne dla tych klas są postaci:

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

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

Równanie granicy decyzyjnej:

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

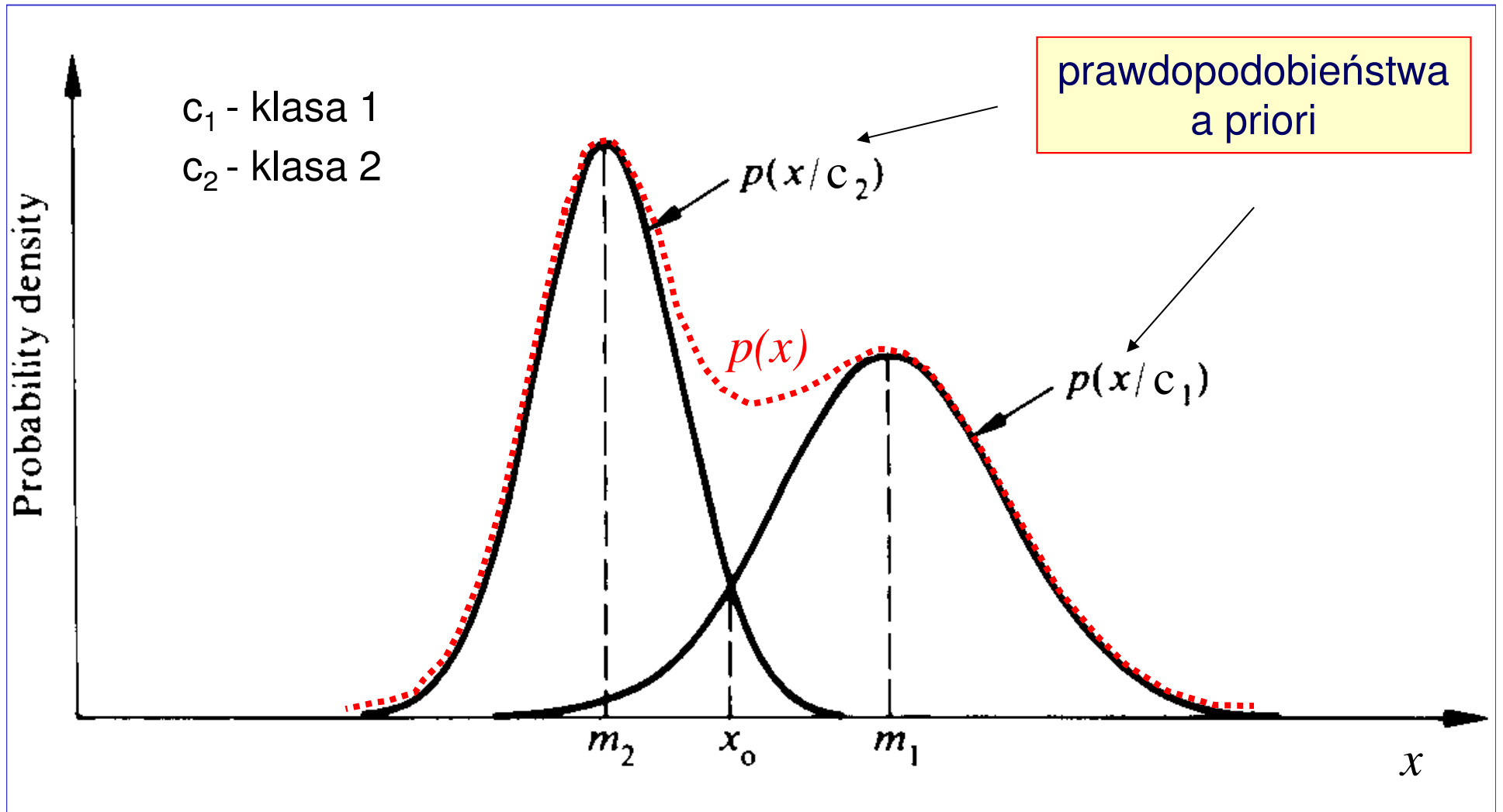
Nowy wektor cech jest klasyfikowany zależnie od znaku $d_{12}(\mathbf{x})$.

Przykład 2: Klasyfikacja znaków do numeracji czeków

Do klasyfikacji tych znaków zastosowano klasyfikator minimalnoodległościowy.

Próbki sygnałów odpowiadające każdemu ze znaków, tworzą zbiór 14 wektorów prototypowych. Klasyfikacja znaków polega na znalezieniu prototypu klasy najbliższego wektorowi cech utworzonego z rozpoznawanego znaku.

Klasyfikator Bayesa dla rozkładów Gaussa



Reguła Bayesa

Z podstaw rachunku prawdopodobieństwa wiadomo, że:

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

zatem:

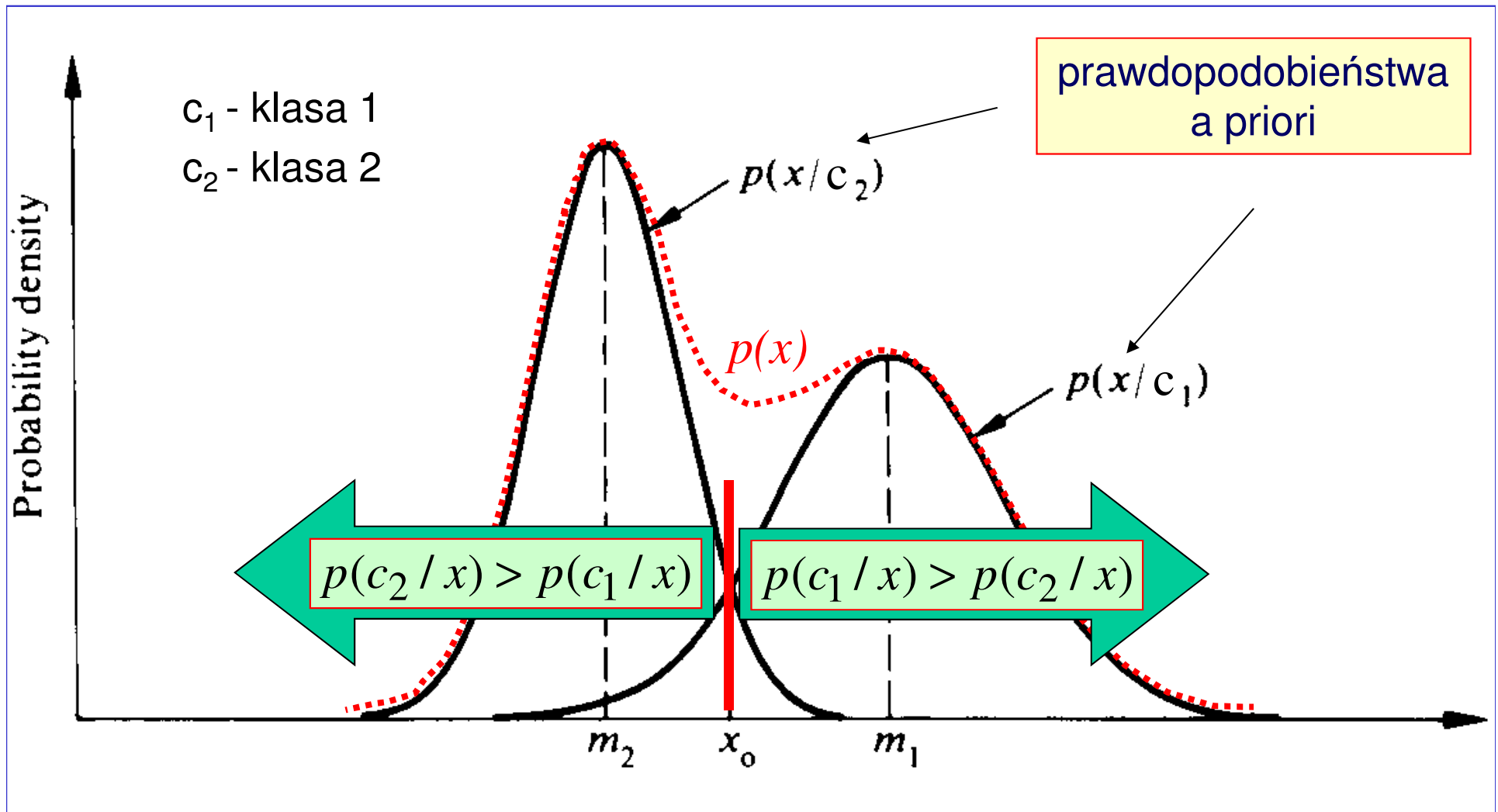
$$p(c_i / \mathbf{x}) = \frac{P(c_i)p(\mathbf{x}' / c_i)}{p(\mathbf{x})}$$

**prawdopodobieństwo
a priori**

**prawdopodobieństwo
a posteriori**

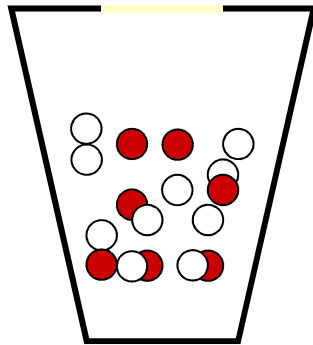
**reguła
Bayesa**

Klasyfikator Bayesa dla rozkładów Gaussa



Klasyfikator Bayesa - przykład

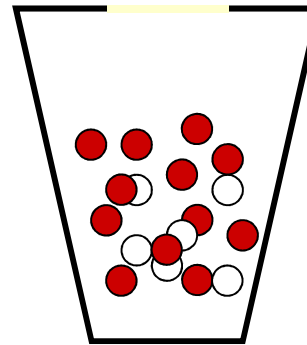
Urna A



20 – czerwone

20 – białe

Urna B



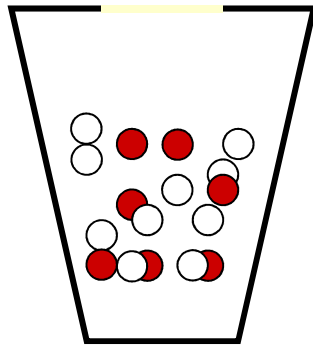
30 – czerwone

10 – białe

Losowo „w ciemno” wybieramy urnę a potem żeton w urnie.
Pytanie: Z której urny wybraliśmy żeton?

Klasyfikator Bayesa - przykład

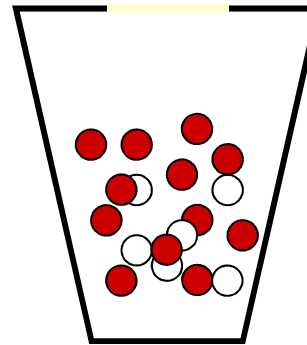
Urna A



20 – czerwone

20 – białe

Urna B



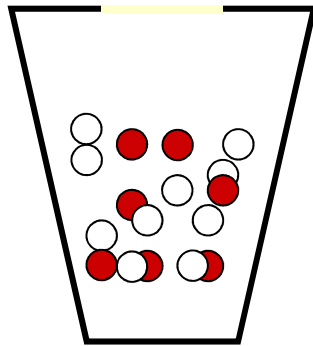
30 – czerwone

10 – białe

Zanim stwierdzimy jaki kolor żetonu wylosowaliśmy mamy podstawy twierdzić (wiedza a priori), że prawdopodobieństwo wybrania każdej z urn wynosi 0.5.

Klasyfikator Bayesa - przykład

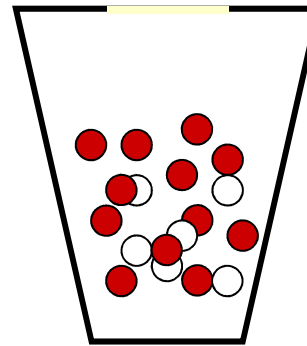
Urna A



20 – czerwone

20 – białe

Urna B



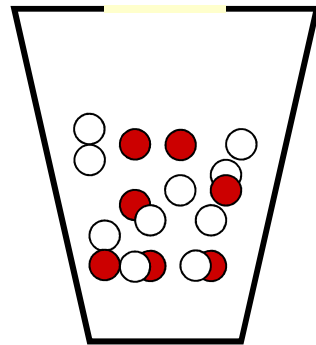
30 – czerwone

10 – białe

Przypuśćmy, że wylosowaliśmy czerwony żeton. Czy możemy teraz zweryfikować naszą hipotezę co do wyboru urny na podstawie nowej wiedzy a posteriori? Tak, odpowiedź mamy na podstawie tw. Bayesa.

Klasyfikator Bayesa - przykład

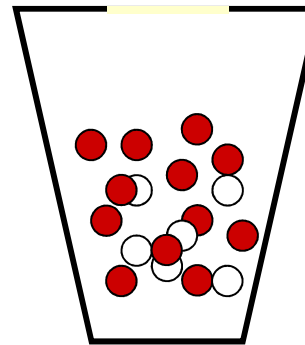
Urna A



20 – czerwone

20 – białe

Urna B



30 – czerwone

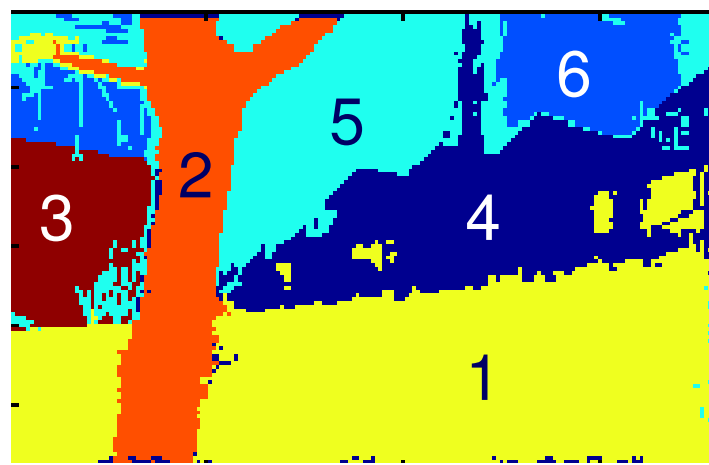
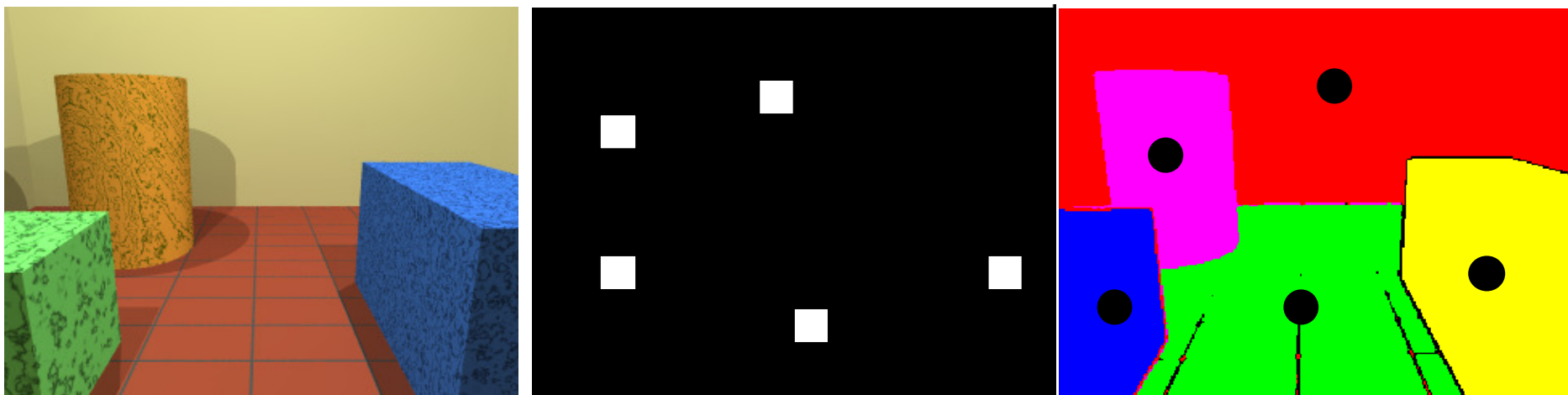
10 – białe

 To ta
urna!
najprawdopodobniej


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

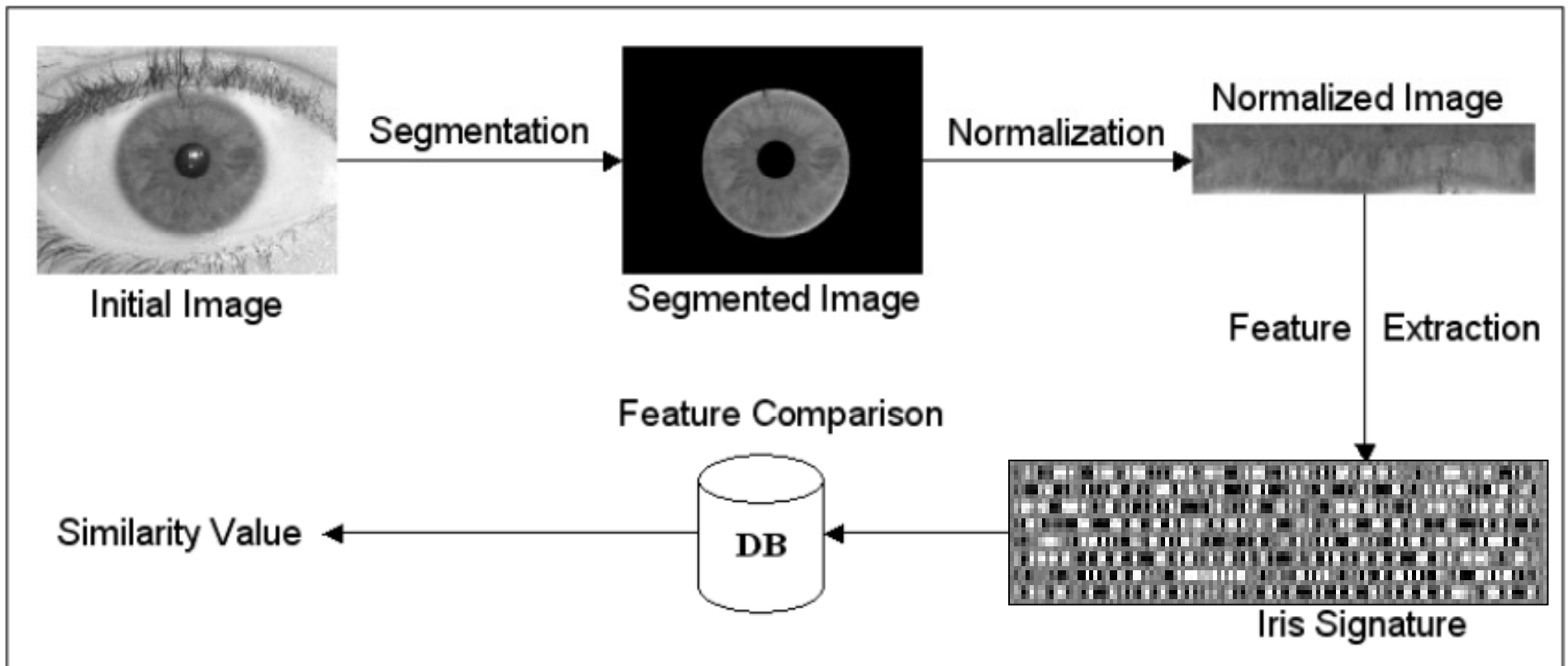
$$p(B / \text{red}) = \frac{P(B)p(\text{red} / B)}{P(A)p(\text{red} / A) + P(B)p(\text{red} / B)} = \frac{0.5 \times 0.75}{0.5 \times 0.5 + 0.5 \times 0.75} = 0.6$$

Klasyfikacja obrazów (tzw. segmentacja)



Paweł Strumiłło, Instytut Elektroniki Politechniki Łódzkiej

Przykłady zadań rozpoznawania obrazów: rozpoznawanie tęczówki oka (biometria)



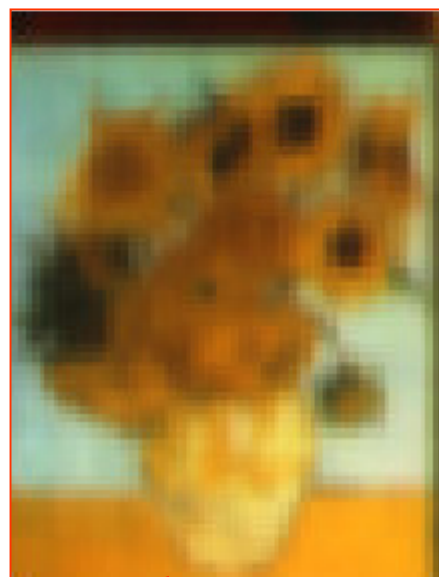
Patentowany algorytm Daugmana

Przykłady zadań rozpoznawania obrazów: Bazy obrazów

„Idea” obrazu
poszukiwanego lub



kopia obrazu
poszukiwanego



Obraz
odszukany



DWT



*C.E. Jacobs, A. Finkelstein, D.H. Salesis,
„Fast multiresolution image quering”, 1999*

Paweł Strumiłło, Instytut Elektroniki Politechniki Łódzkiej