KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

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

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

dr inż. Artur Klepaczko

Eksploracja danych
Klasyfikatory liniowe

Zadanie nr 13 – Studia podyplomowe „Przetwarzanie i analiza obrazów biomedycznych"

Politechnika Łódzka
Instytut Elektroniki

90-924 Łódź, ul. Żeromskiego 116,
tel. 042 631 28 83
www.kapitalludzki.p.lodz.pl

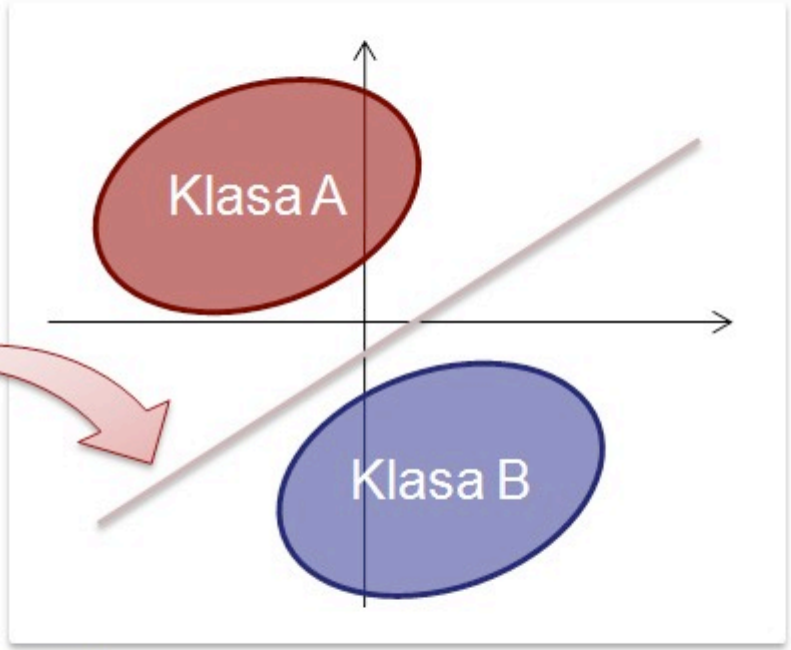
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Dyskryminacja liniowa

Klasy separowalne liniowo

Funkcja dyskryminacyjna (postać ogólna)

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

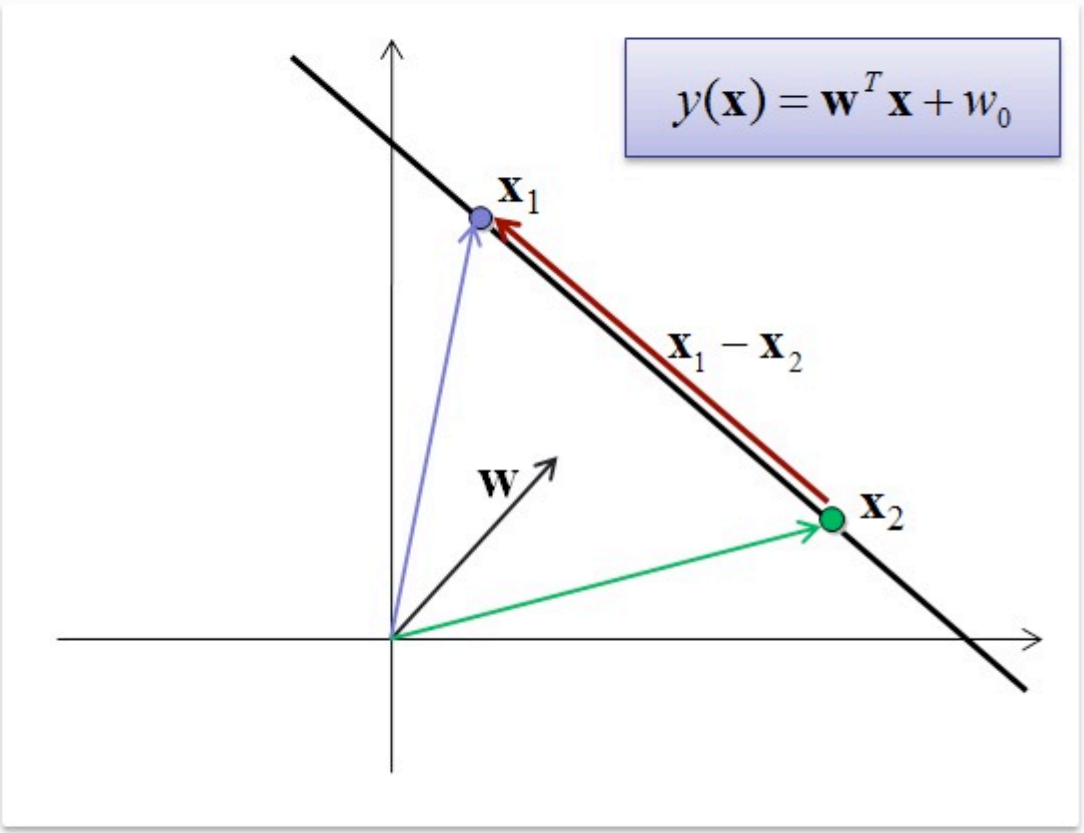


Przestrzeń dwuwymiarowa	– linia
Przestrzeń trójwymiarowa	– płaszczyzna
Przestrzeń n-wymiarowa	– hiperpłaszczyzna

$$y(\mathbf{x}) \geq 0 \quad \Rightarrow \quad \mathbf{x} \in A$$
$$y(\mathbf{x}) < 0 \quad \Rightarrow \quad \mathbf{x} \in B$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Funkcja liniowa — wektor wag



$$y(\mathbf{x}_1) = y(\mathbf{x}_2) = 0$$

$$\mathbf{w}^T\mathbf{x}_1 + w_0 = \mathbf{w}^T\mathbf{x}_2 + w_0 \Rightarrow$$
$$\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) = 0$$

Iloczyn skalarny wektorów o niezerowej długości jest równy 0 tylko wtedy, gdy wektory te są ortogonalne (prostopadłe).

Wektor wag określa orientację hiperpłaszczyzny decyzyjnej.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

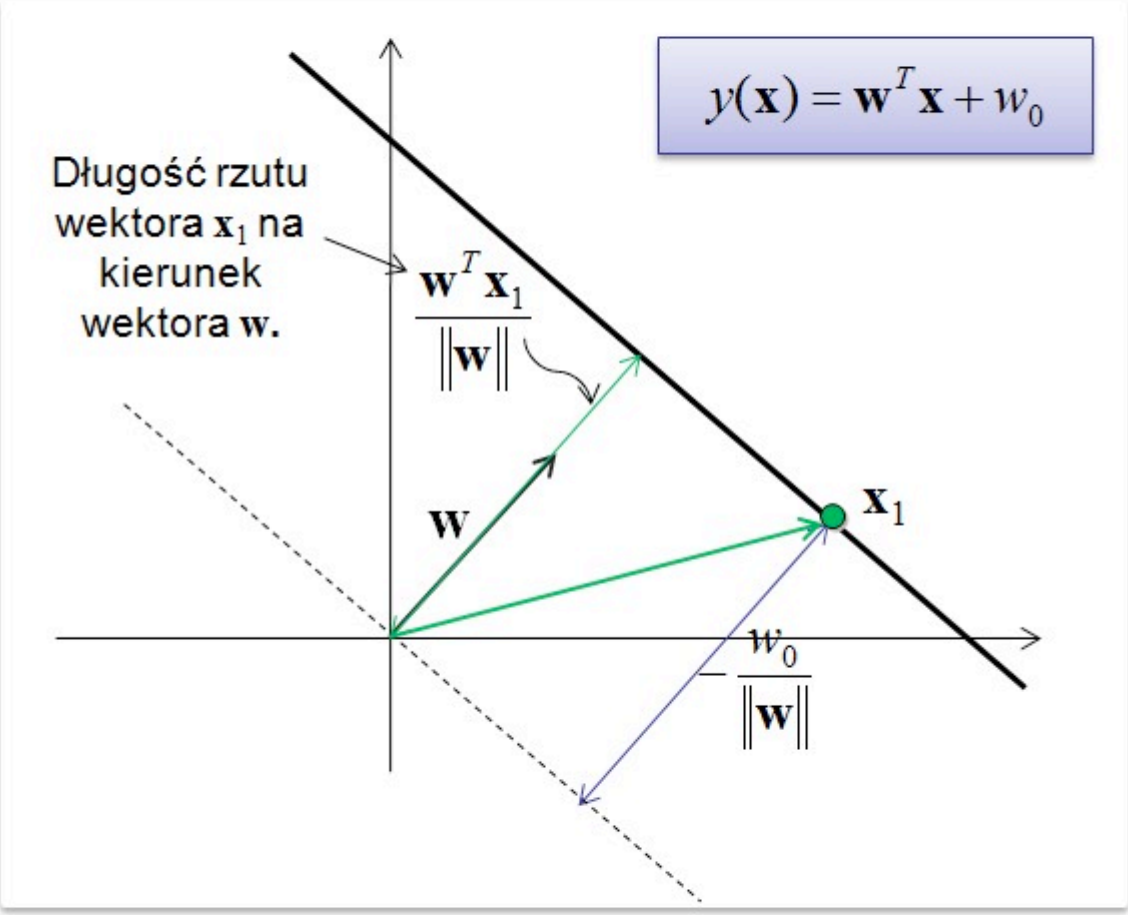
Funkcja liniowa — odchylenie

$$y(\mathbf{x}_1) = 0$$

$$\mathbf{w}^T\mathbf{x}_1 + w_0 = 0 \Rightarrow$$

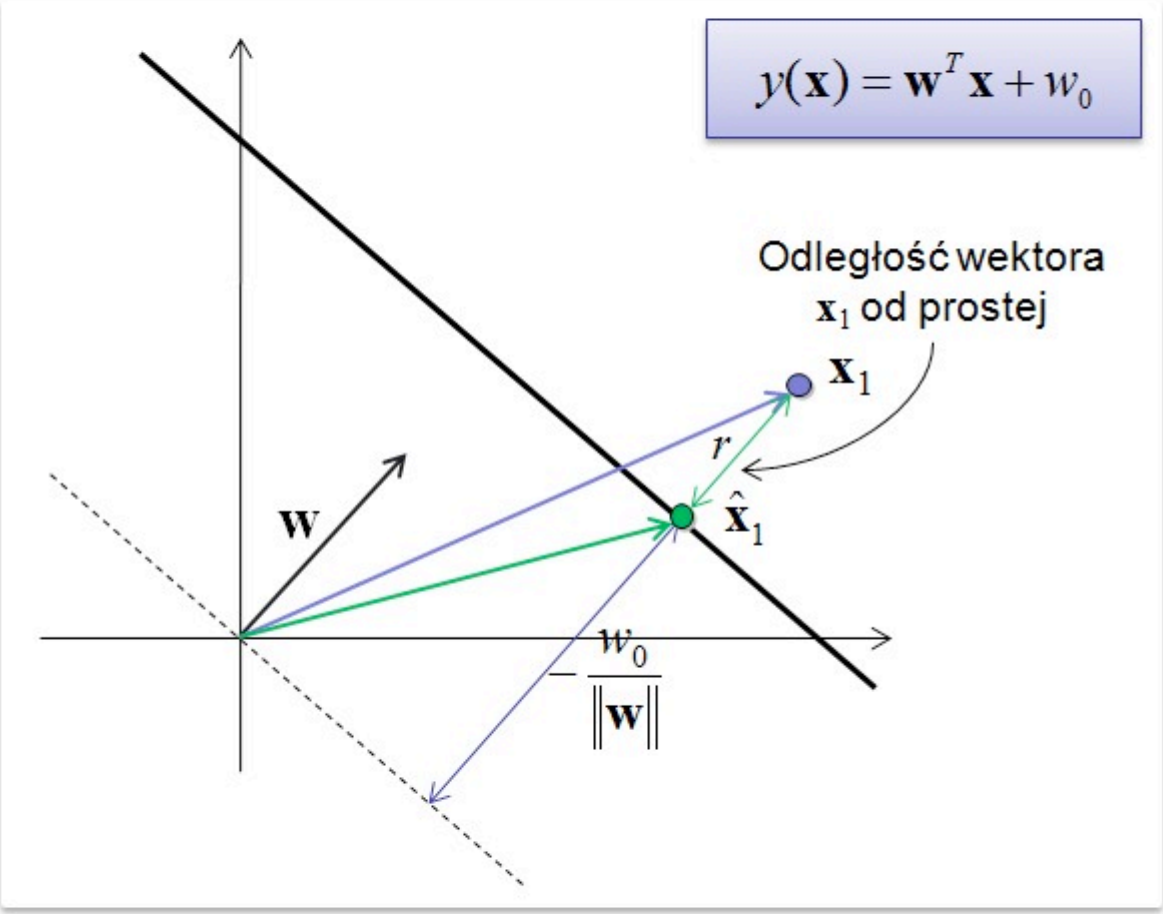
$$\frac{\mathbf{w}^T\mathbf{x}_1}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|}$$

Parametr odchylenia określa położenie hiperpłaszczyzny decyzyjnej.



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Odległość punktu od linii prostej



$$\mathbf{x}_1 = \hat{\mathbf{x}}_1 + r\frac{\mathbf{w}}{\|\mathbf{w}\|}$$

$$\mathbf{w}^T\mathbf{x}_1 = \mathbf{w}^T\hat{\mathbf{x}}_1 + r\frac{\mathbf{w}^T\mathbf{w}}{\|\mathbf{w}\|}$$

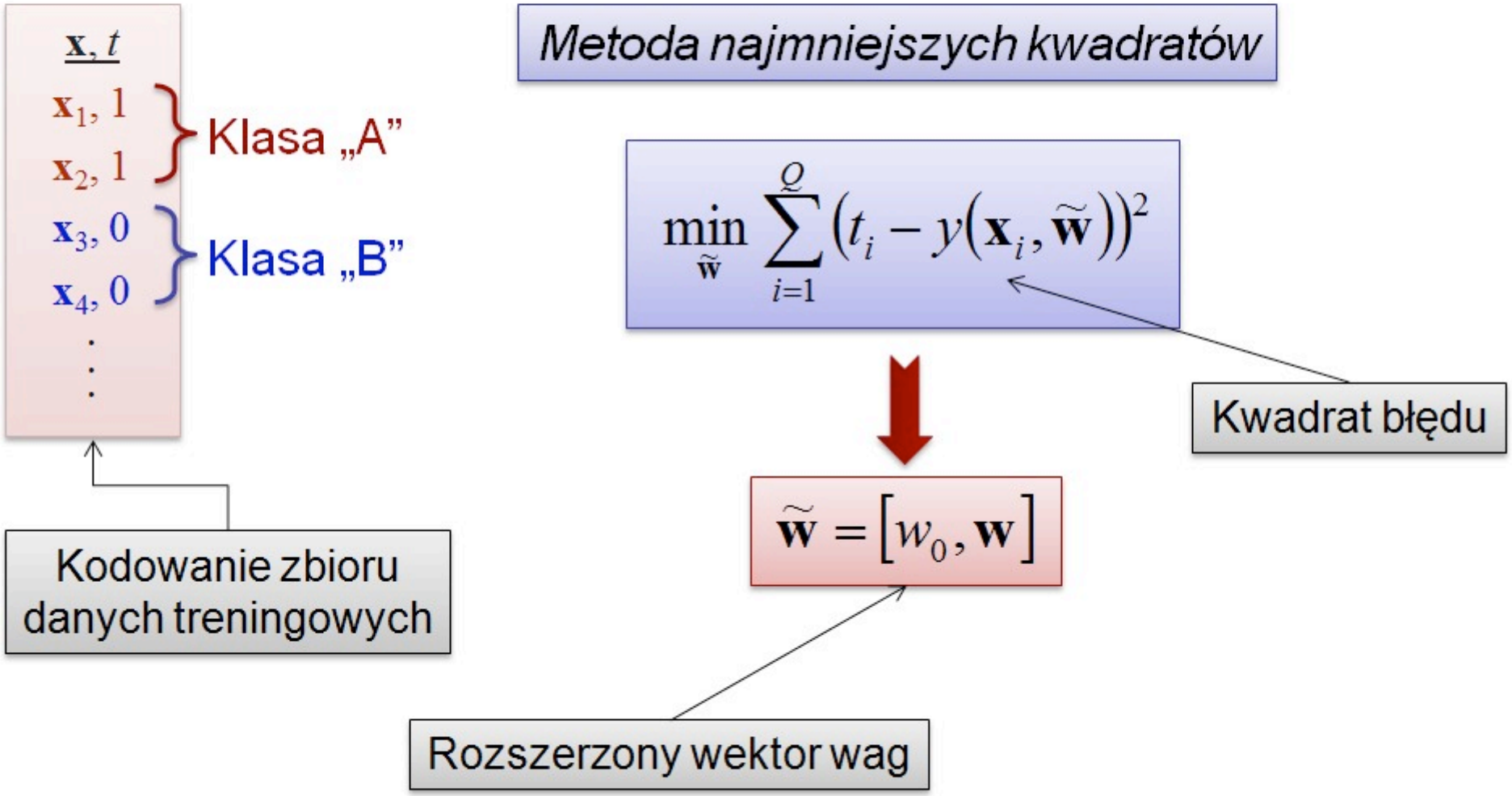
$$\mathbf{w}^T\mathbf{x}_1 + w_0 = \mathbf{w}^T\hat{\mathbf{x}}_1 + w_0 + r\frac{\mathbf{w}^T\mathbf{w}}{\|\mathbf{w}\|}$$

$$y(\mathbf{x}_1) = r\frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|}$$

$$r = \frac{y(\mathbf{x}_1)}{\|\mathbf{w}\|}$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

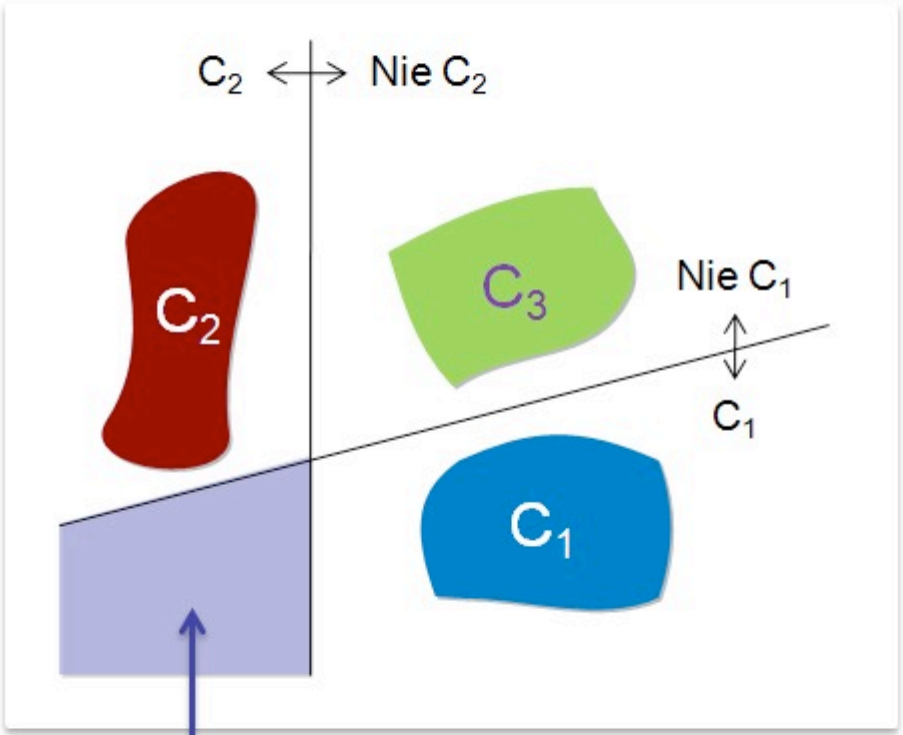
Uczenie klasyfikatora — regresja liniowa



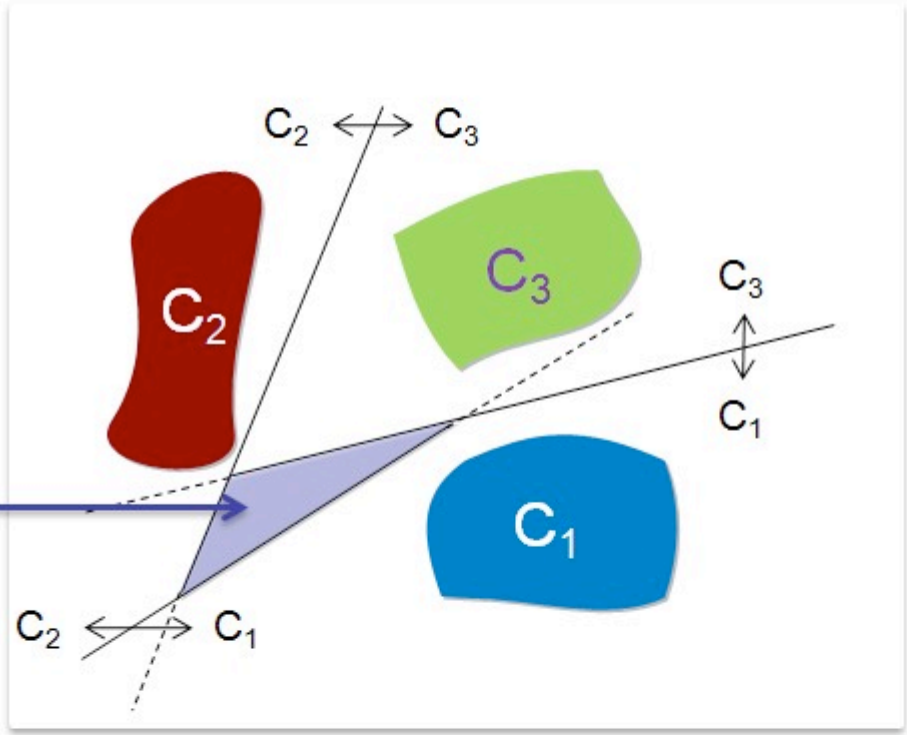
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Rozwinięcie do przypadku wielu klas

1. Zestaw K–1 linii decyzyjnych



2. Zestaw $K(K–1)/2$ linii decyzyjnych

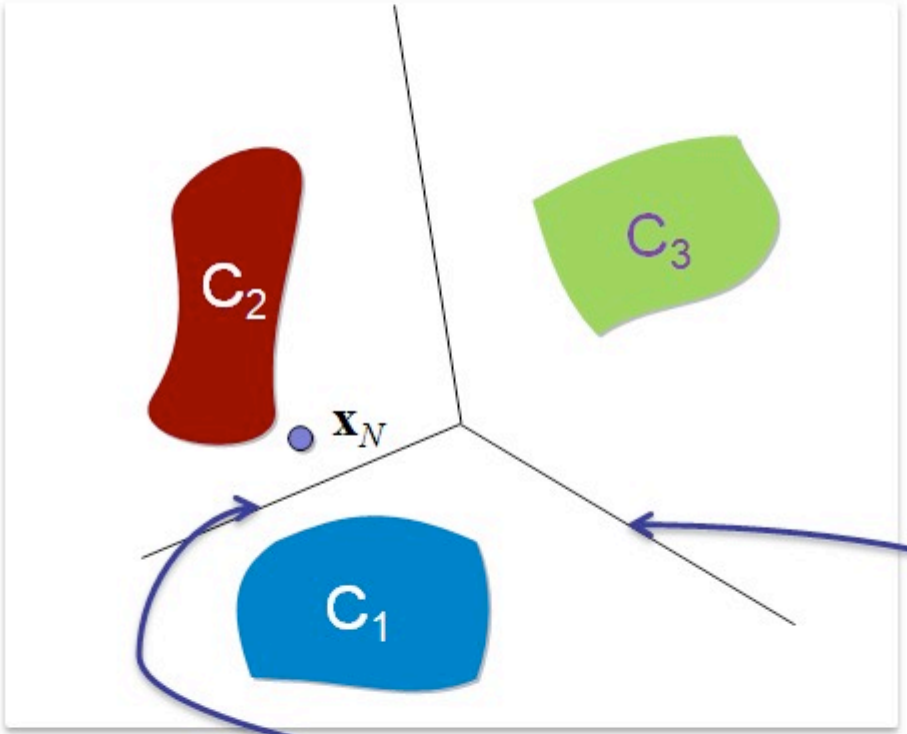


Obszary niejednoznaczne

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Rozwiązanie dla problemu wielu klas

3. Zestaw *K* linii decyzyjnych



$y(\mathbf{x})$ — funkcja przynależności wektora do danej klasy

Np.

$y_1(\mathbf{x}_N) = 0.3$

$y_2(\mathbf{x}_N) = 0.6$

$y_3(\mathbf{x}_N) = 0.1$

Wektor $\mathbf{x}_N$ klasyfikujemy do C_2

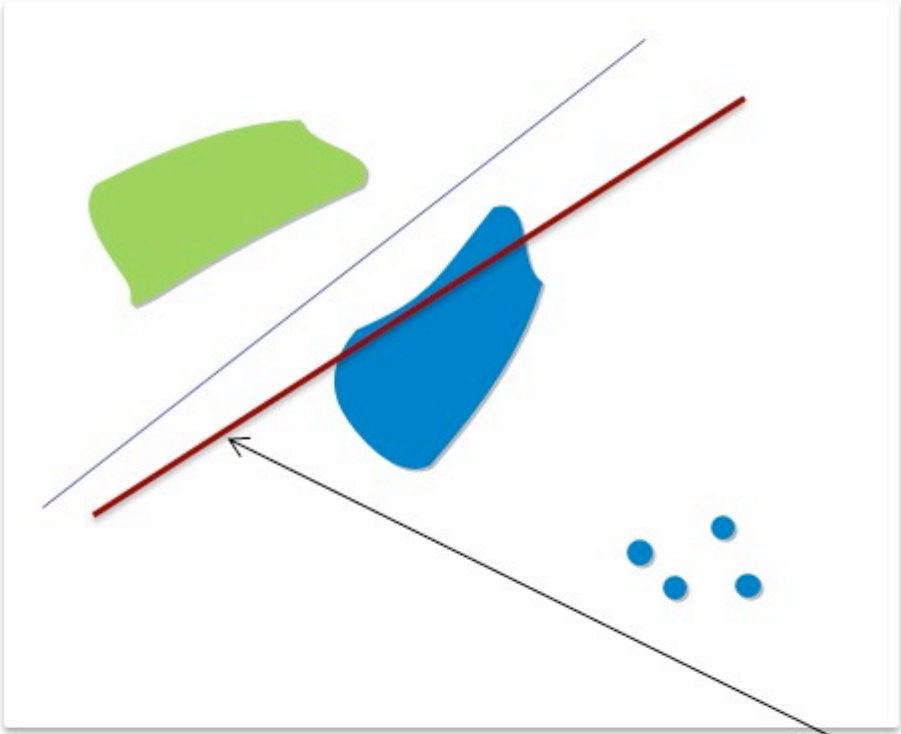
$y_1(\mathbf{x}) - y_3(\mathbf{x})$

$y_1(\mathbf{x}) - y_2(\mathbf{x})$

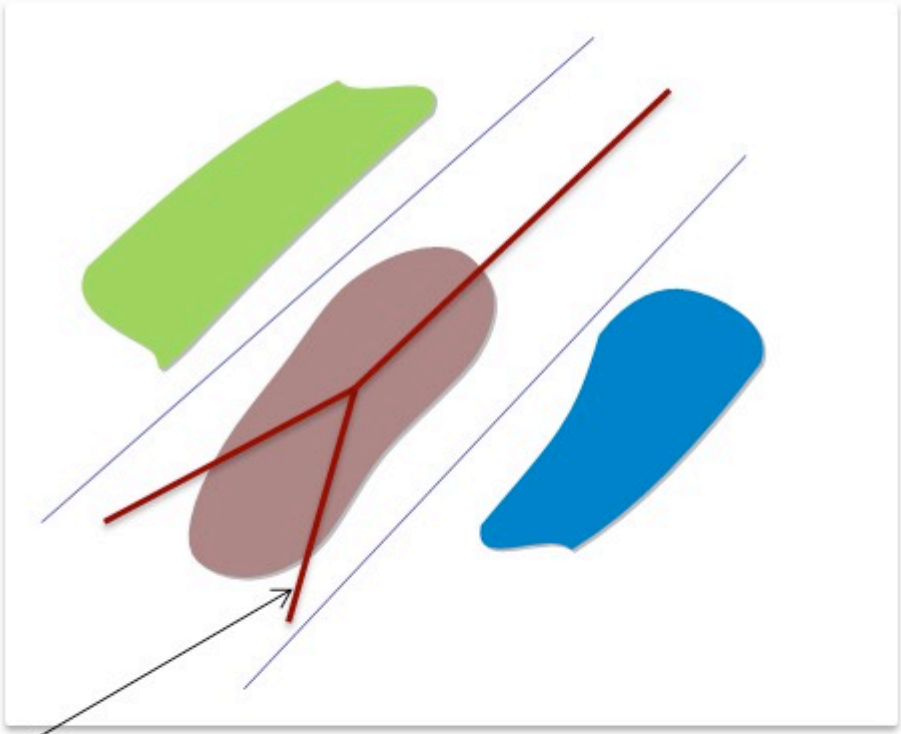
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Regresja liniowa — ograniczenia

Wrażliwość na wartości oddalone (ang. *outliers)*



Słaba separacja w przypadku wielu klas



Linie decyzyjne wyznaczone metodą najmniejszych kwadratów.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Regresja a szacowanie prawdopodobieństwa

Regresja liniowa

$$\Pr[y=1|\mathbf{x}] = w_0 + \mathbf{w}^T\mathbf{x}$$

Prawdopodobieństwo przynależności do klasy zakodowanej jako y=1.

Jednakże, wartość funkcji decyzyjnej wykracza poza zakres (0,1).

Transformacja zmiennej docelowej

$$\Pr[y=1|\mathbf{x}] \rightarrow \ln\frac{\Pr[y=1|\mathbf{x}]}{1-\Pr[y=1|\mathbf{x}]}$$

Funkcja *logitowa*

Regresja logistyczna

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Regresja logistyczna

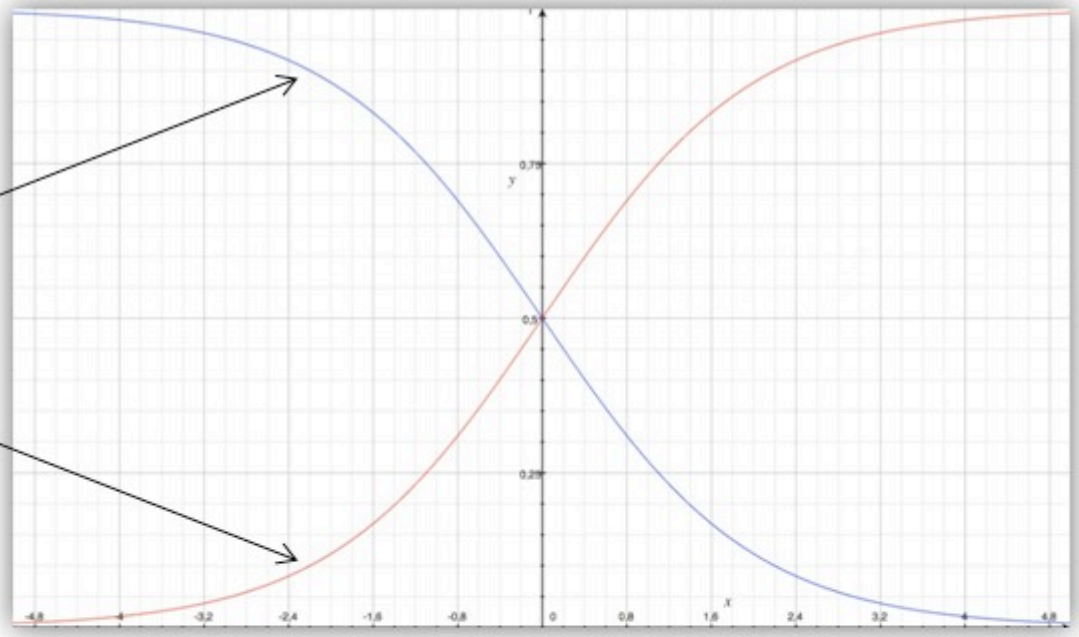
$$\ln\frac{\Pr[y=0\,|\,\mathbf{x}]}{1-\Pr[y=0\,|\,\mathbf{x}]}=w_0+\mathbf{w}^T\mathbf{x}$$

Aproksymacja zmiennej docelowej określonej funkcją logitową przy użyciu funkcji liniowej.

Rozwikłanie modelu

$$\Pr[y=0\,|\,\mathbf{x}]=\frac{\exp(w_0+\mathbf{w}^T\mathbf{x})}{1+\exp(w_0+\mathbf{w}^T\mathbf{x})}$$

$$\Pr[y=1\,|\,\mathbf{x}]=\frac{1}{1+\exp(w_0+\mathbf{w}^T\mathbf{x})}$$



Wartości obu oszacowań mieszczą się zawsze w zakresie (0,1)

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Metoda największej wiarygodności

Funkcja wiarygodności

$$LL = \sum_{i=1}^{Q} (1 - t_i) \log(1 - \Pr[1 \mid \mathbf{x}_i]) + t_i \log(\Pr[1 \mid \mathbf{x}_i])$$

Maksymalizacja

Parametry modelu logistycznego → $\mathbf{w}$, w_0

Reguła klasyfikacyjna

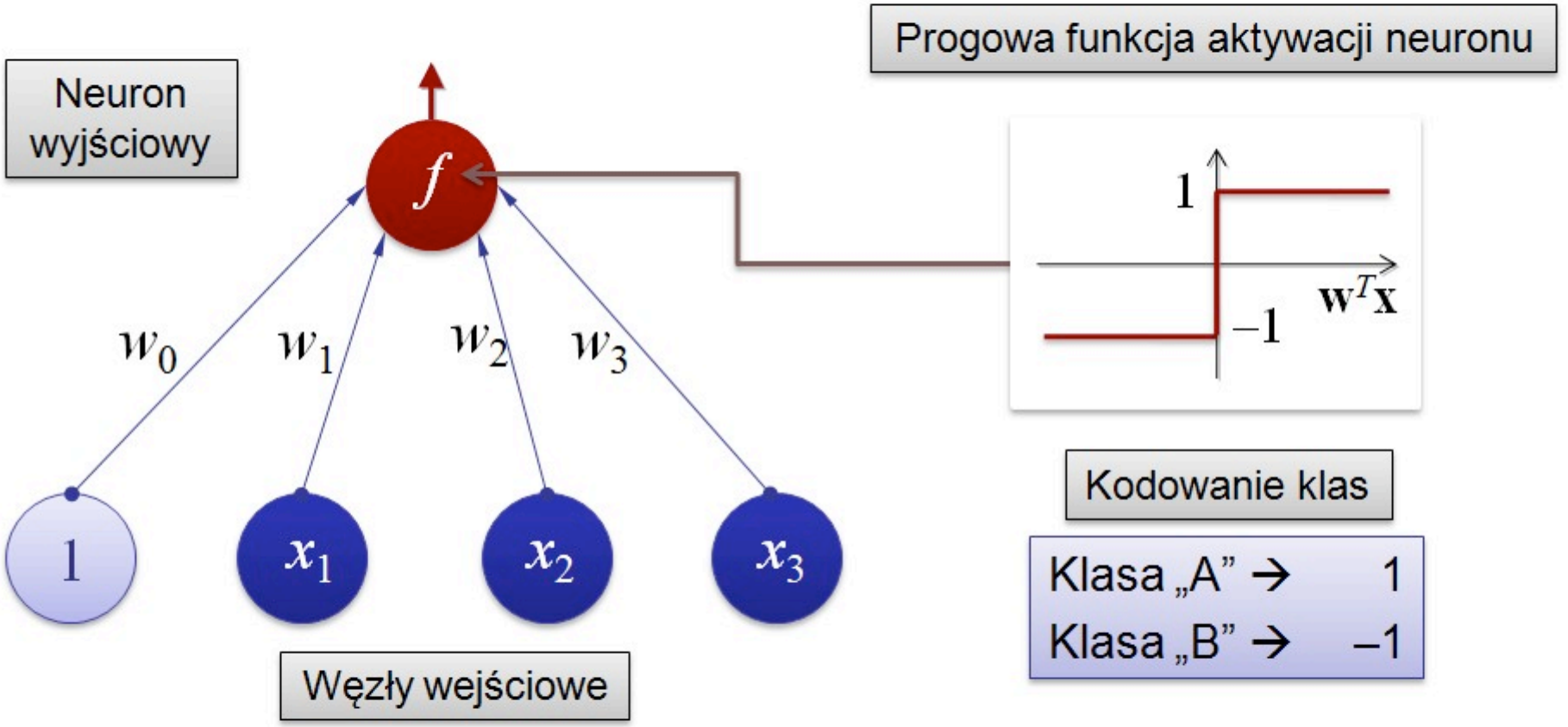
Jeśli

$$\Pr[1 \mid \mathbf{x}_N] > \Pr[0 \mid \mathbf{x}_N]$$

to wektor $\mathbf{x}_N$ zaliczamy do klasy zakodowanej jako 1.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

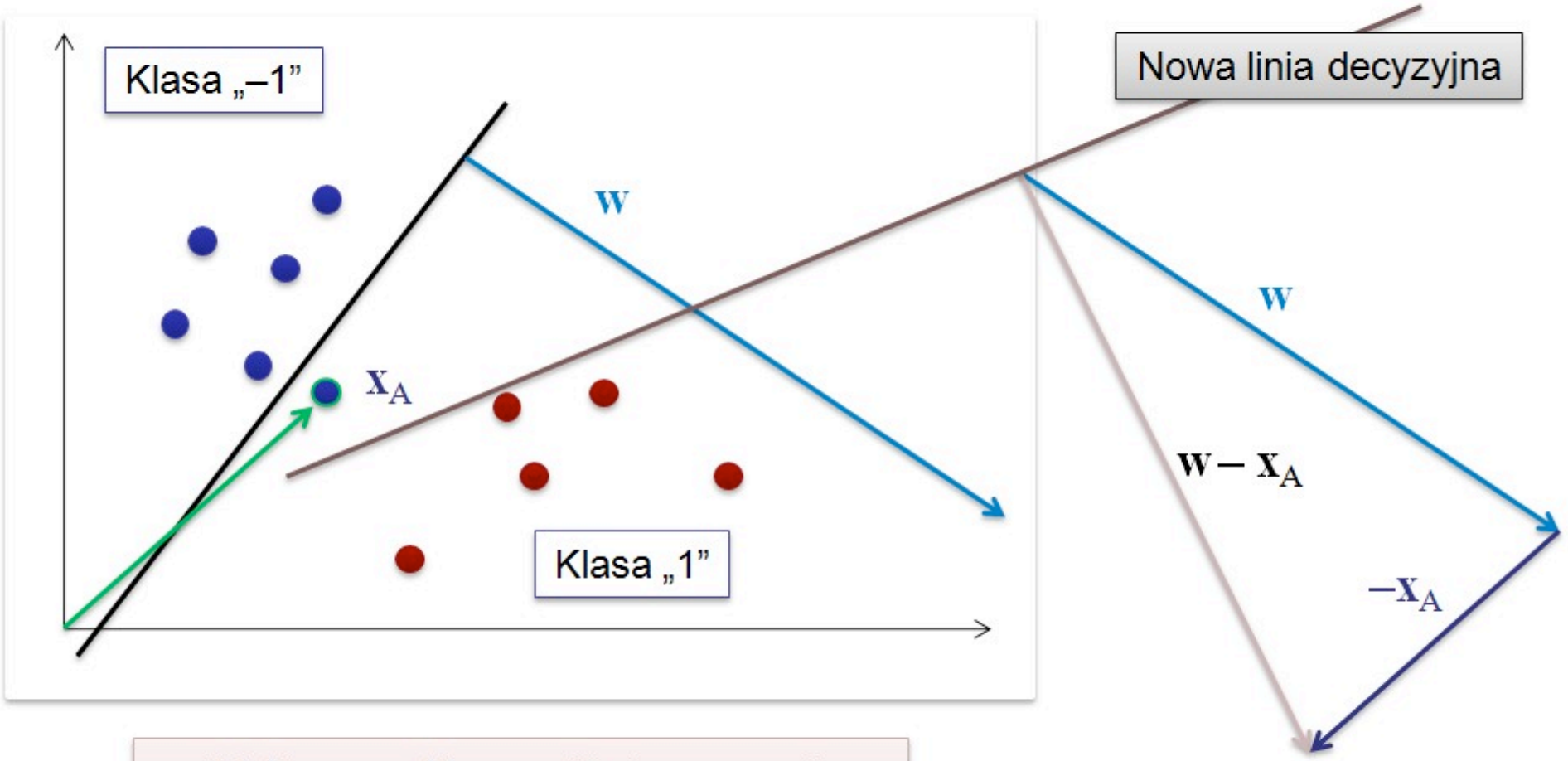
Perceptron



Regresja — estymowanie wartości prawdopodobieństwa przynależności do danej klasy
Reguła perceptronowa — „po prostu” wyznaczanie hiperpłaszczyzny decyzyjnej

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

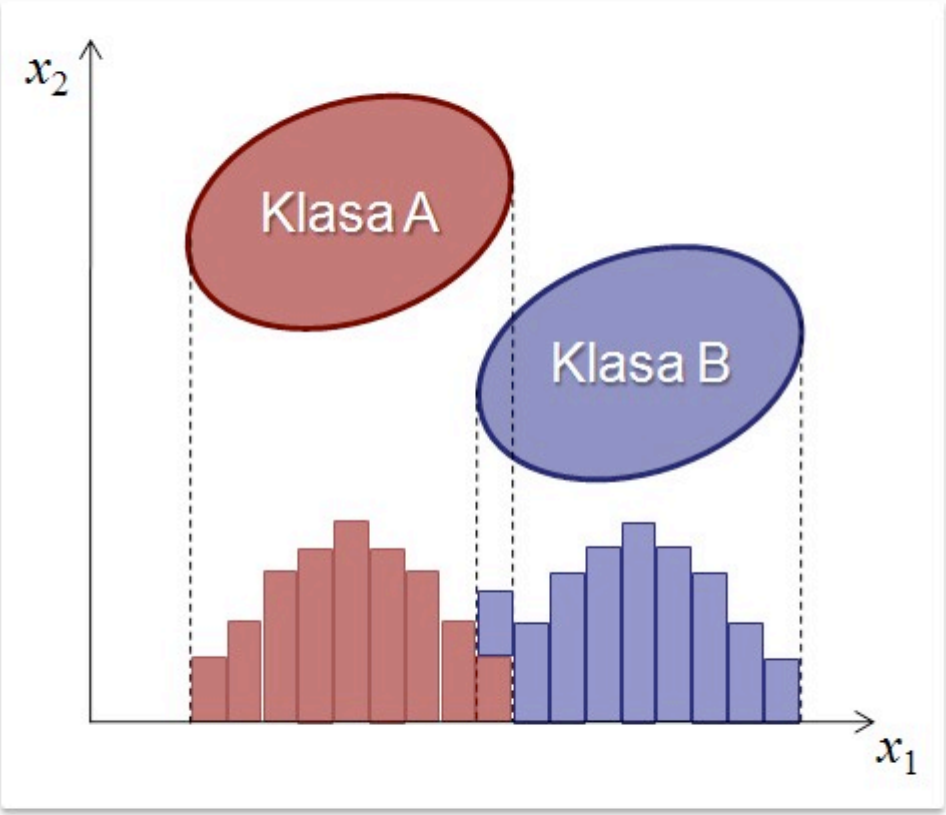
Uczenie perceptronu — modyfikacja wag

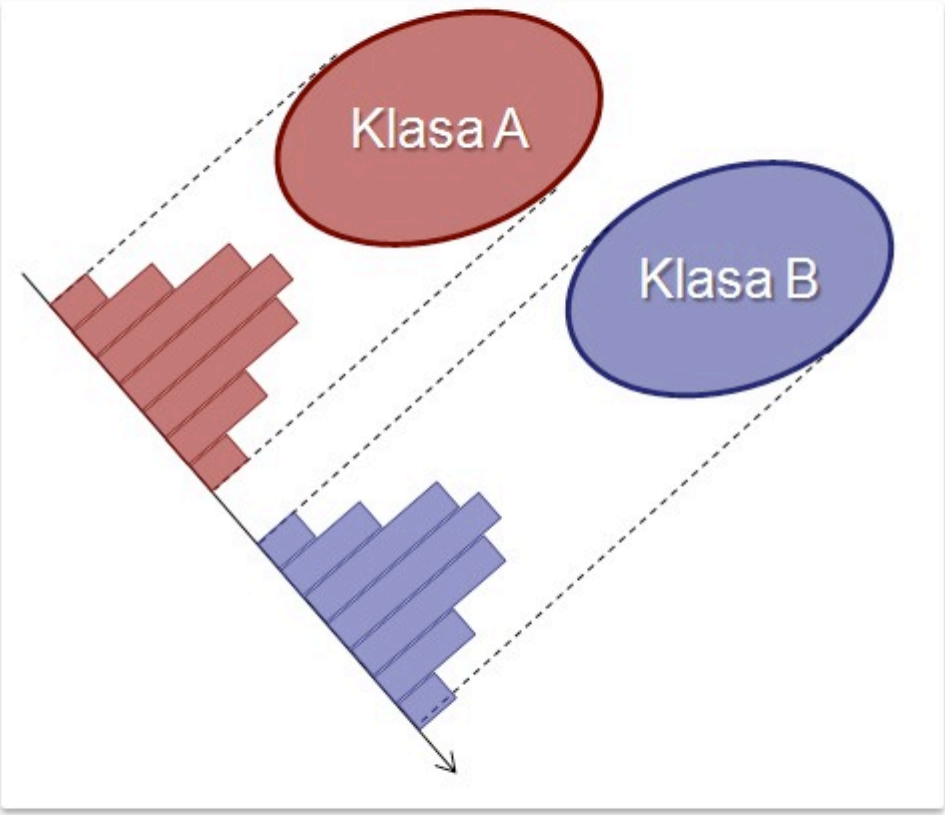


Wektor $\mathbf{x}_A$ z klasy „–1” niepoprawnie zaklasyfikowany do klasy „1”

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Liniowa analiza dyskryminacyjna





Zadanie liniowej analizy dyskryminacyjnej (LDA) polega na znalezieniu w przestrzeniu cech kierunku zapewniającego najlepszą możliwą separację klas.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zmienność międzygrupowa

Wektor średni

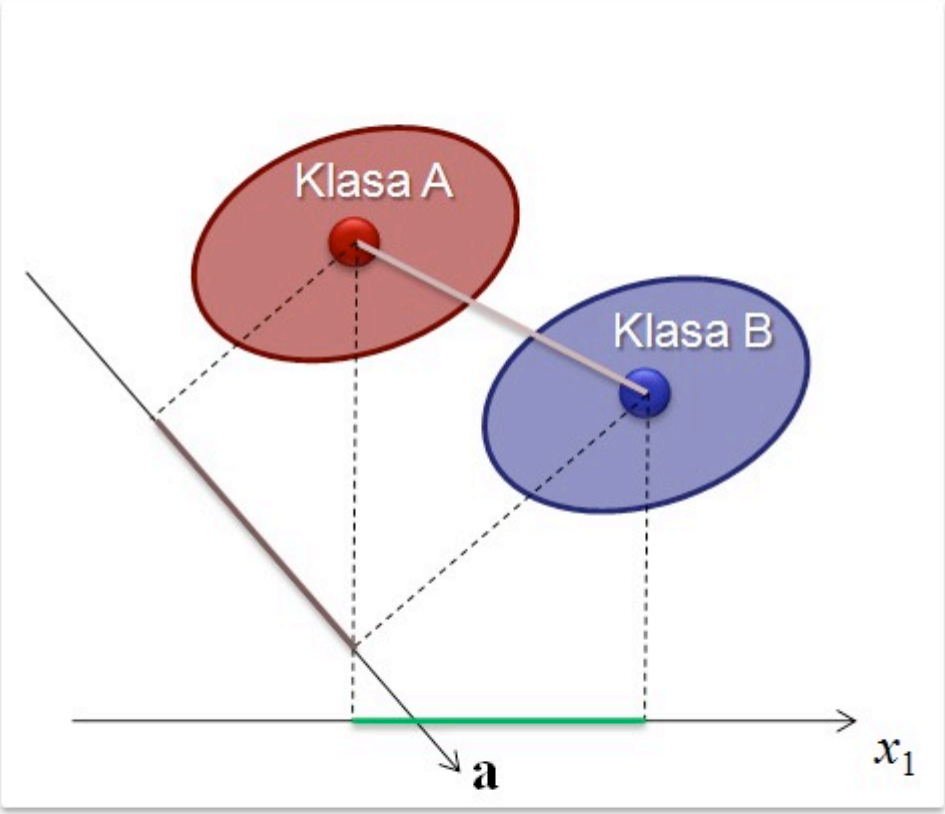
$$\bar{\mathbf{x}} = \left[\bar{x}^{(1)}, \bar{x}^{(2)}, \ldots, \bar{x}^{(j)}, \ldots \bar{x}^{(N)}\right]$$

$$\bar{x}^{(j)} = \frac{1}{Q}\sum_{i=1}^{Q} x_i^{(j)}$$

Miara zmienności międzygrupowej

$$\left(\mathbf{a}^T\bar{\mathbf{x}}_A - \mathbf{a}^T\bar{\mathbf{x}}_B\right)^2$$

Odległość średnich wektorów klas wzdłuż kierunku **a**



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zmienność wewnątrzgrupowa

Macierz kowariancji

$$\mathbf{S}_k = \frac{1}{Q_k - 1} \sum_{i=1}^{Q_k} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

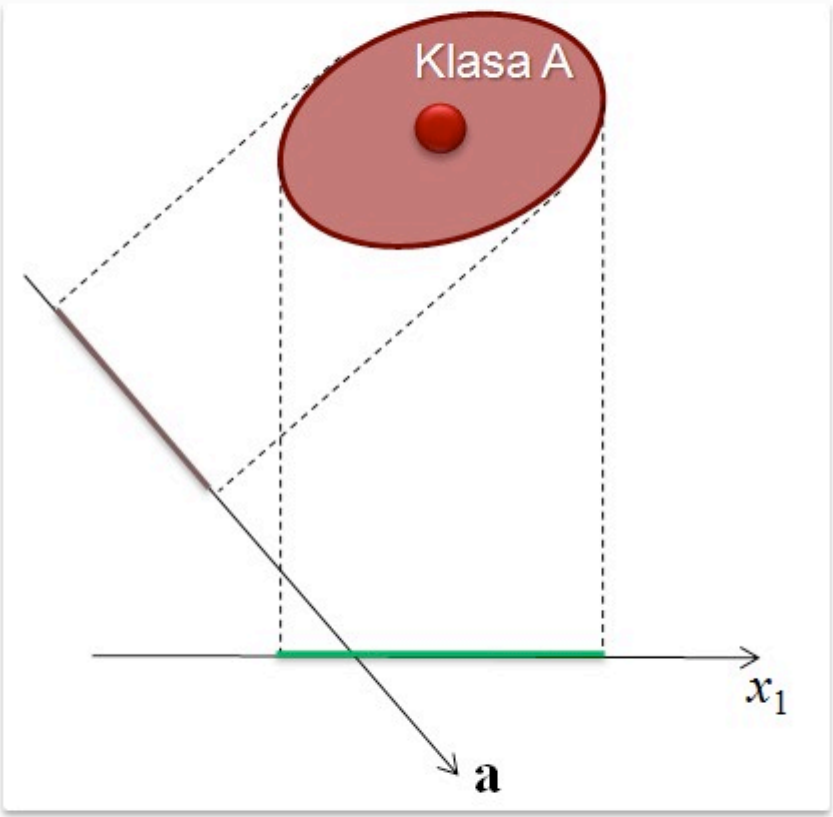
Pojedyncza klasa

$$\mathbf{W} = \frac{1}{Q - 2} \sum_{k=1}^{2} (Q_k - 1)\mathbf{S}_k$$

Wspólna dla obu klas

Miara zmienności wewnątrzgrupowej

$$\mathbf{a}^T \mathbf{W} \mathbf{a}$$



Rozproszenie wektorów wokół średnich wektorów klas w danym kierunku

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Analiza LDA a klasyfikacja

Zadanie LDA

Znaleźć taki kierunek $\hat{\mathbf{a}}$, który maksymalizuje wyrażenie

$$\frac{\left(\mathbf{a}^T\overline{\mathbf{x}}_A - \mathbf{a}^T\overline{\mathbf{x}}_B\right)^2}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

Małe rozproszenie wewnątrz dobrze odseparowanych klas.

Klasyfikacja

Nowy wektor $\mathbf{x}_N$ należy zaklasyfikować do A jeśli

$$\left|\hat{\mathbf{a}}^T\mathbf{x}_N - \hat{\mathbf{a}}^T\overline{\mathbf{x}}_A\right| < \left|\hat{\mathbf{a}}^T\mathbf{x}_N - \hat{\mathbf{a}}^T\overline{\mathbf{x}}_B\right|$$

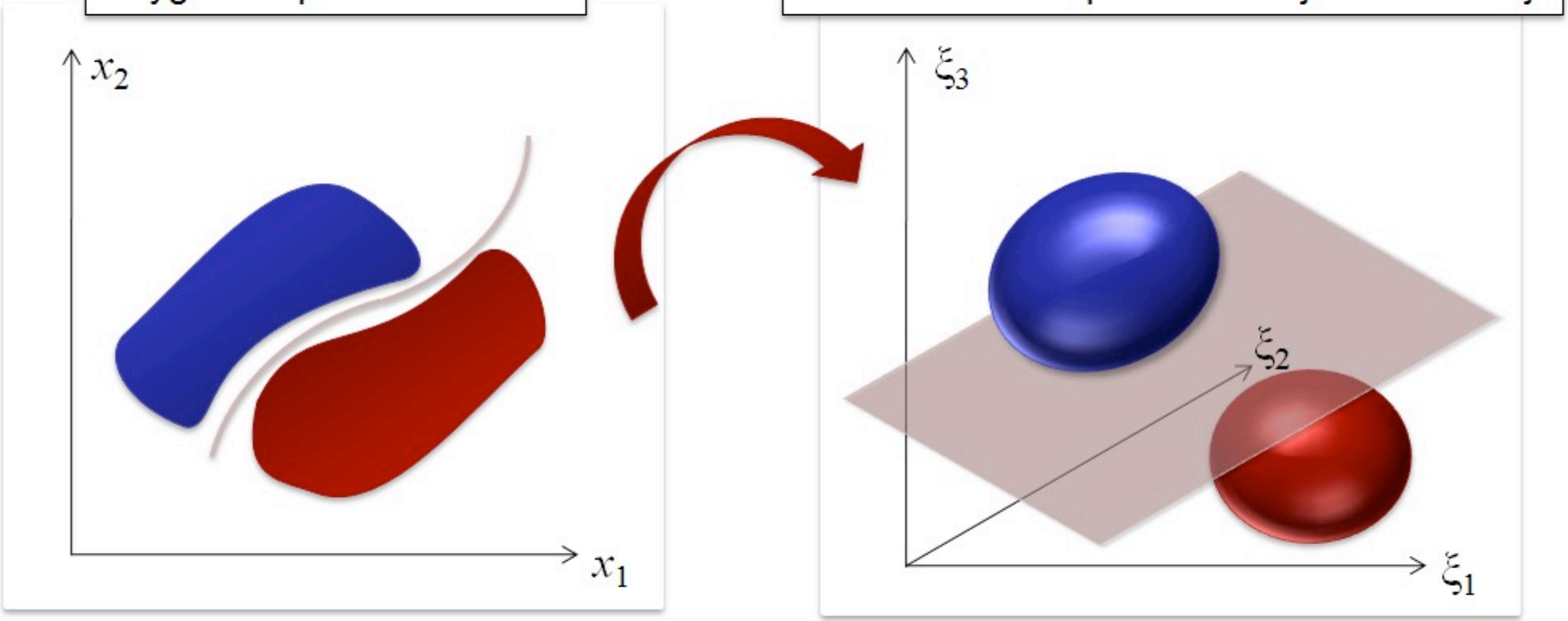
Wektor testowy leży bliżej środka klasy A.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Nieliniowe granice między klasami

Oryginalna przestrzeń cech

Przestrzeń cech po nieliniowej transformacji



Problem nieliniowej separacji klas można rozwiązać za pomocą klasyfikatorów liniowych po zastosowaniu nieliniowej transformacji przestrzeni cech.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Nieliniowa transformacja przestrzeni cech

Przykładowa transformacja przestrzeni cech

$$\xi_1 = x_1^2 x_2 \quad \xi_2 = x_1 x_2^2 \quad \xi_3 = x_1^3$$

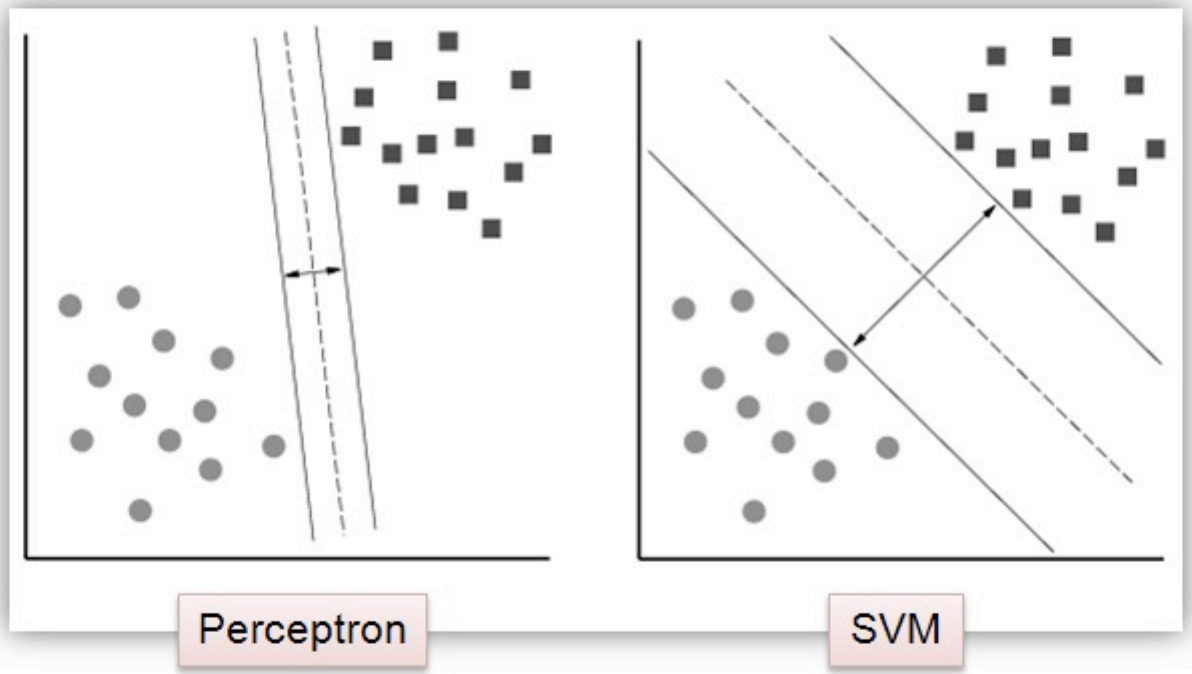
Często, aby uzyskać poprawną klasyfikację wymiar nowej przestrzeni cech musi być znacznie większy niż wymiar przestrzeni oryginalnej.

- duża złożoność obliczeniowa, zwłaszcza dla dużych wymiarów oryginalnej przestrzeni atrybutów
- niebezpieczeństwo nadmiernego dopasowania

Odpowiedzią na te problemy jest algorytm wektorów podpierających (ang. *support vector machines*)

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

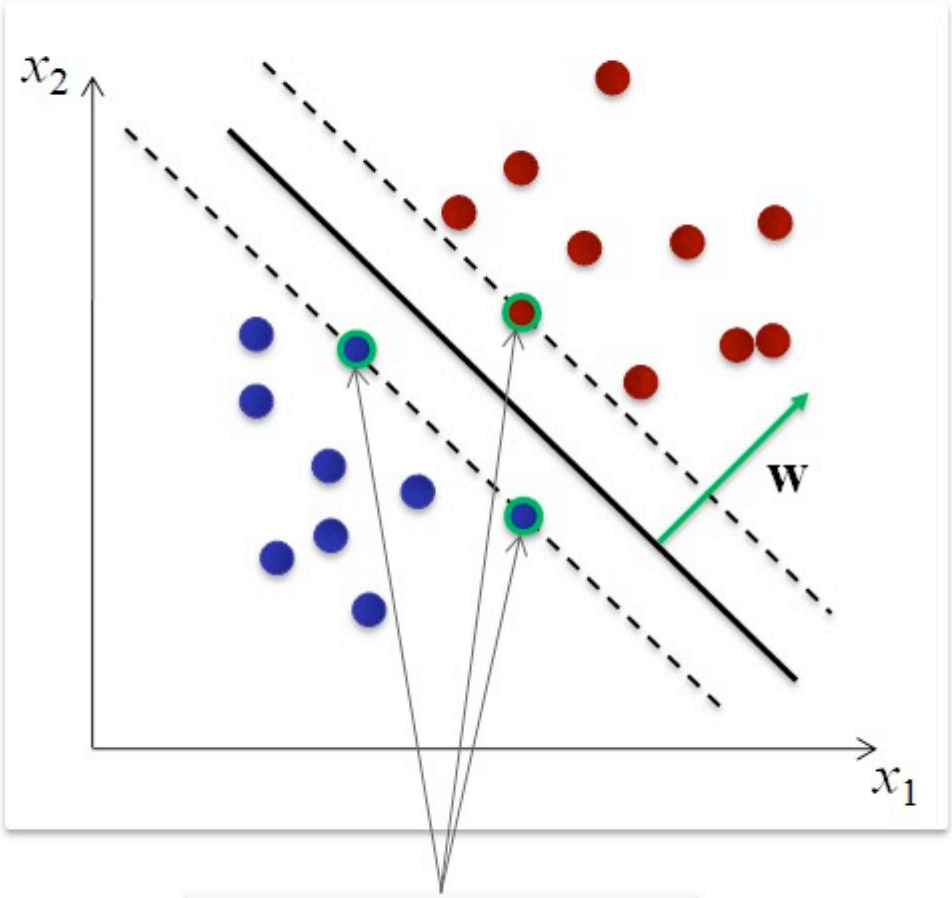
Algorytm *Support Vector Machines*



W algorytmie SVM konstruowana jest hiperpłaszczyzna decyzyjna w kierunku ortogonalnym do kierunku największego marginesu rozdzielającego klasy (ang. *optimal margin hyperplane*).

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Konstrukcja optymalnej hiperpłaszczyzny



Wektory podpierające

Równanie hiperpłaszczyzny

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

$$y(\mathbf{x}) = b + \sum_{i-\text{indeks wektora podpierającego}} \alpha_i t_i \mathbf{x}_i^T \mathbf{x}$$

Współczynniki α_i znajdowane poprzez rozwiązanie zadania kwadratowej optymalizacja z ograniczeniami… (!)

Na szczęście istnieją gotowe implementacje → *sequential minimal optimization* (SMO)

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Funkcje jądra

Transformacja wektorów danych

$$\mathbf{x} \to \Phi(\mathbf{x})$$

Iloczyn skalarny w oryginalnej przestrzeni cech

$$\mathbf{x}_i^T \cdot \mathbf{x}_j$$

Iloczyn skalarny w przestrzeni transformowanej

$$\Phi^T(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \quad \Leftrightarrow \quad K(\mathbf{x}_i, \mathbf{x}_j)$$

K – funkcja jądra (ang. *kernel function)*

Wielomianowe funkcje jądra

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \cdot \mathbf{x}_j)^n$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \cdot \mathbf{x}_j + 1)^n$$

Radialna funkcja jądra

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{c}\right)$$

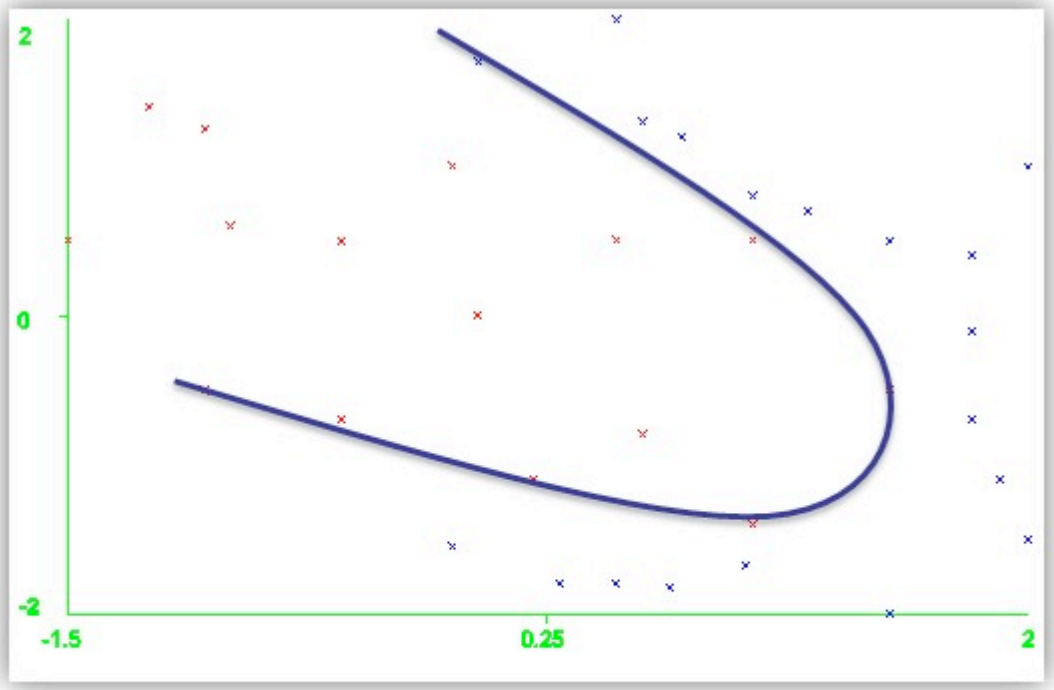
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Algorytm SVM dla przypadku klas nieliniowo separowalnych

Hiperpowierzchnia decyzjna

$$y(\mathbf{x}) = b + \sum_{i} \alpha_i t_i K(\mathbf{x}_i, \mathbf{x})$$

i – indeks wektora podpierającego



Zalety algorytm SVM

- Granica decyzyjna konstruowana jest w oparciu o niewielką liczbę wektorów podpierających
  - małe ryzyko nadmiernego dopasowania
  - odporność na wektory oddalone (ang. *outliers*)

- Dzięki zastosowaniu funkcji jądra możliwe jest rozwiązywania problemów nieliniowych bez zwiększania stopnia złożoności obliczeniowej.

dr inż. Artur Klepaczko

Eksploracja danych
Ocena klasyfikacji

Zadanie nr 13 – Studia podyplomowe „Przetwarzanie i analiza obrazów biomedycznych”

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

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

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

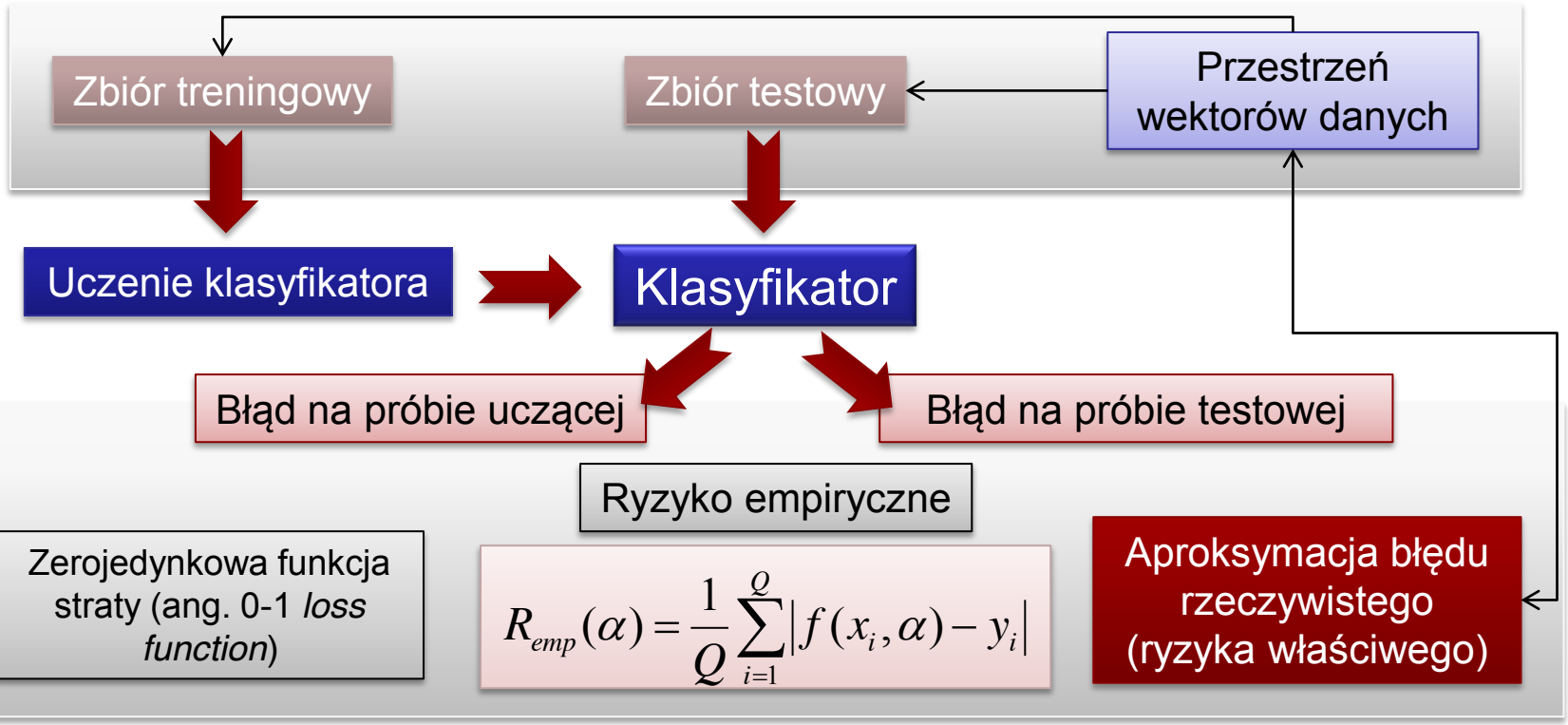
Politechnika Łódzka
Instytut Elektroniki

90-924 Łódź, ul. Żeromskiego 116,
tel. 042 631 28 83
www.kapitalludzki.p.lodz.pl

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

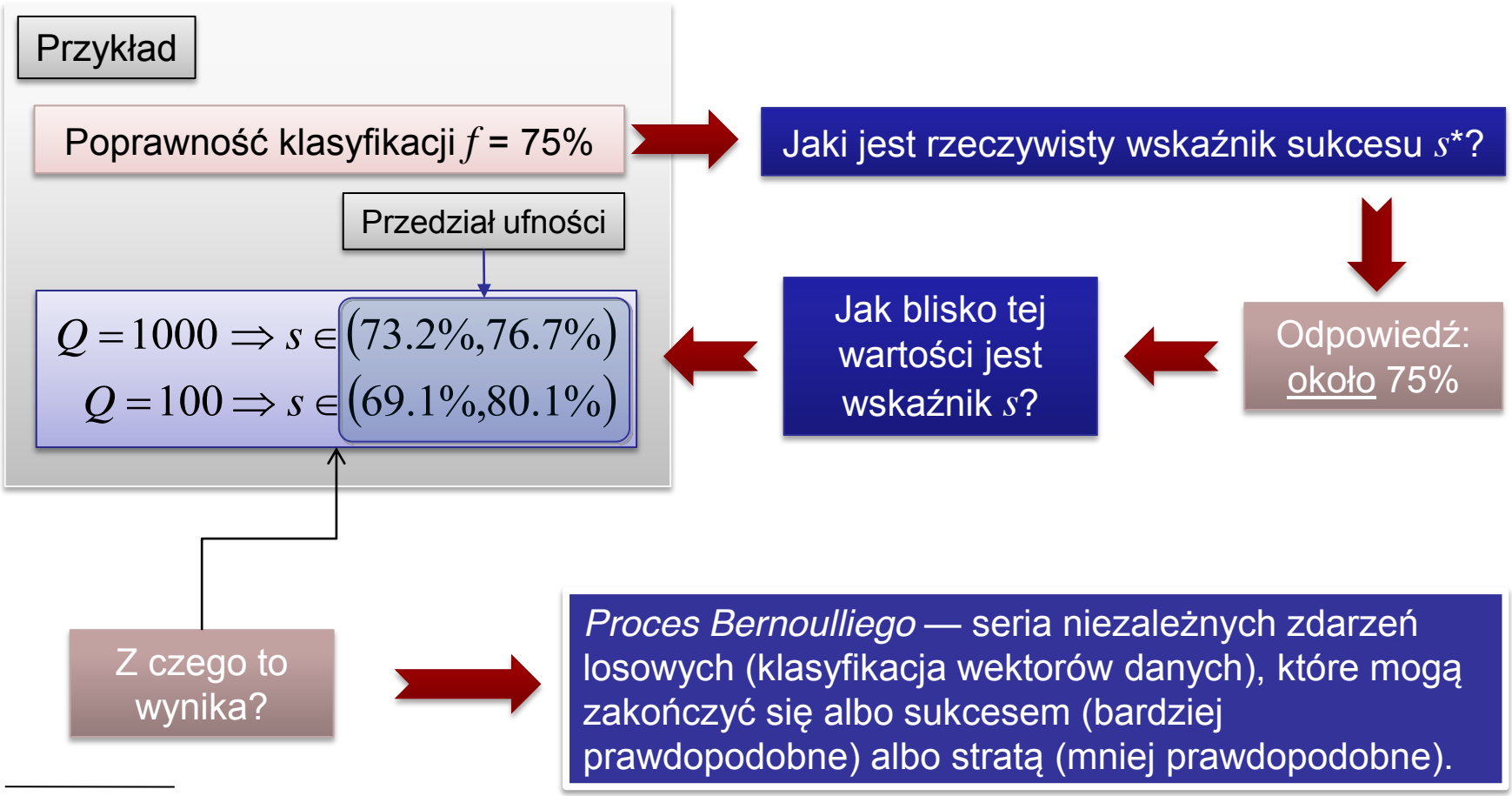
Ocena algorytmu klasyfikacji

Na ile algorytm nauczony na podstawie próby uczącej będzie zdolny do prawidłowej predykcji klasy nowych wektorów danych?



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przedział ufności



*Wskaźnik sukcesu = 1 – błąd klasyfikacji

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

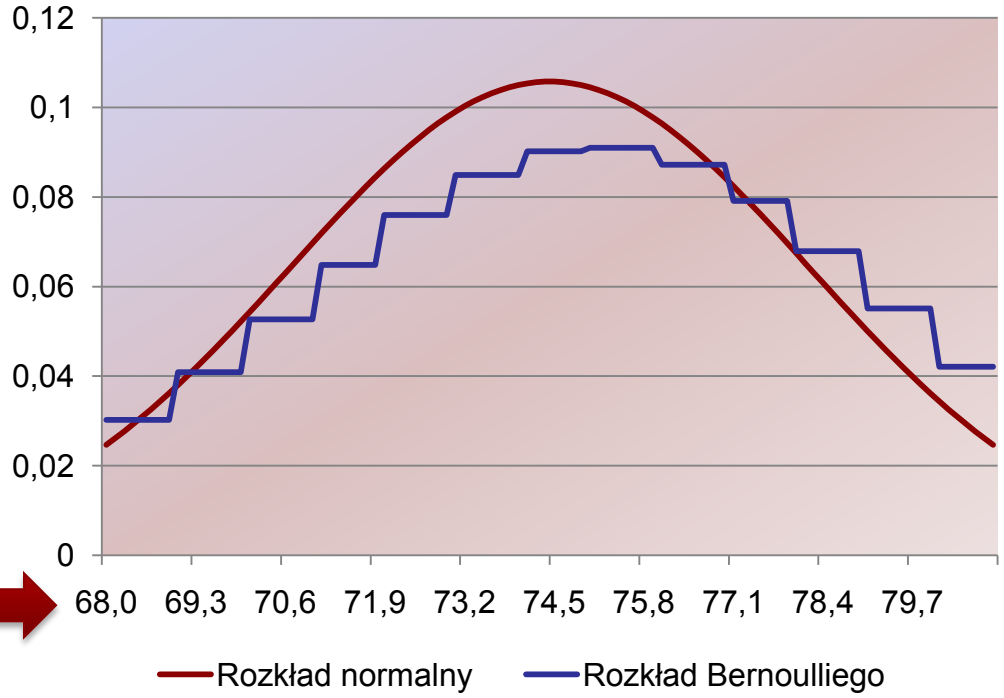
Proces Bernoulliego a rozkład normalny

Średnia szansa na to, aby pojedyncza próba należąca do procesu Bernoulliego zakończyła się sukcesem wynosi s.

Średnia liczba sukcesów dla serii prób Bernoulliego wynosi również s.

W przypadku dużej serii zdarzeń, proces Bernoulliego zbliża się do rozkładu Gaussa.

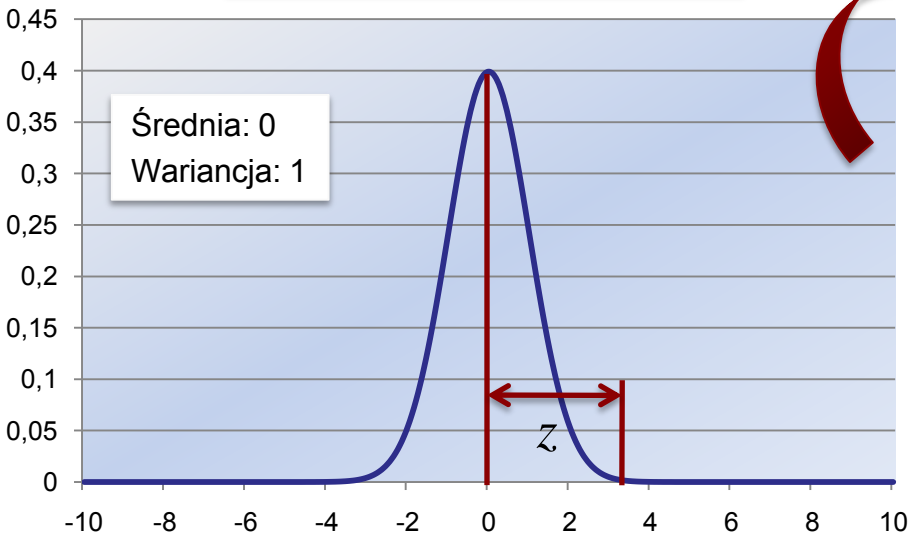
Średnia poprawność klasyfikacji [%] dla różnych serii zdarzeń



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Obliczanie przedziału ufności

Rozkład normalny $N(0,1)$



Granice przedziałów ufności

$\Pr[X \geq z]$	z
0,1%	3,09
0,5%	2,58
1%	2,33
5%	1,65
10%	1,28
20%	0,84

Wynik klasyfikacji

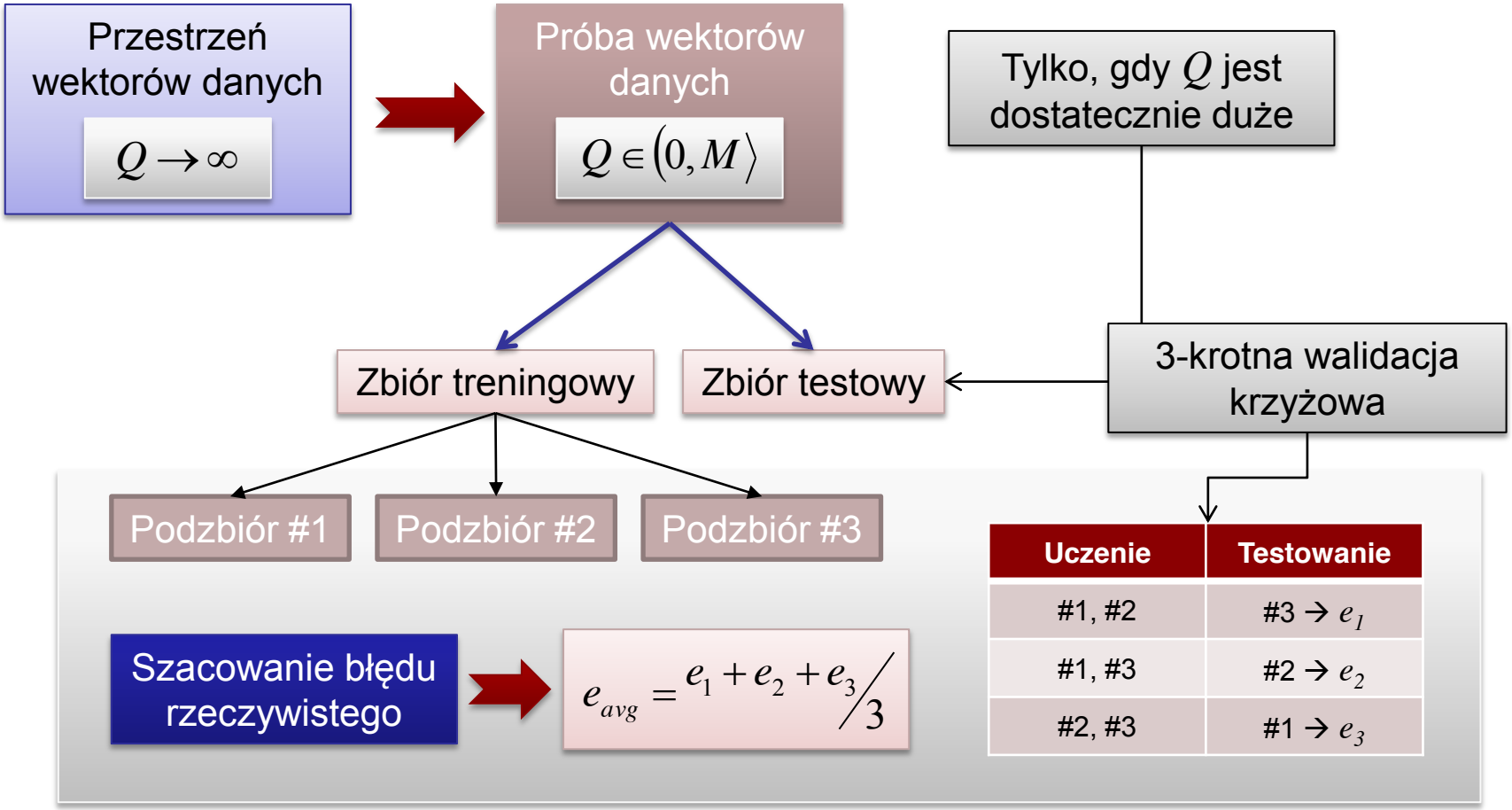
$$s = \left(f + \frac{z^2}{2Q} \pm z\sqrt{\frac{f}{Q} - \frac{f^2}{Q} + \frac{z^2}{4Q^2}} \right) \Big/ \left(1 + \frac{z^2}{Q} \right)$$

Z tabeli

Prawdopodobieństwo tego, że zmienna X przyjmie wartość oddaloną o 2,33 od średniej (powyżej lub poniżej) wynosi 2%.

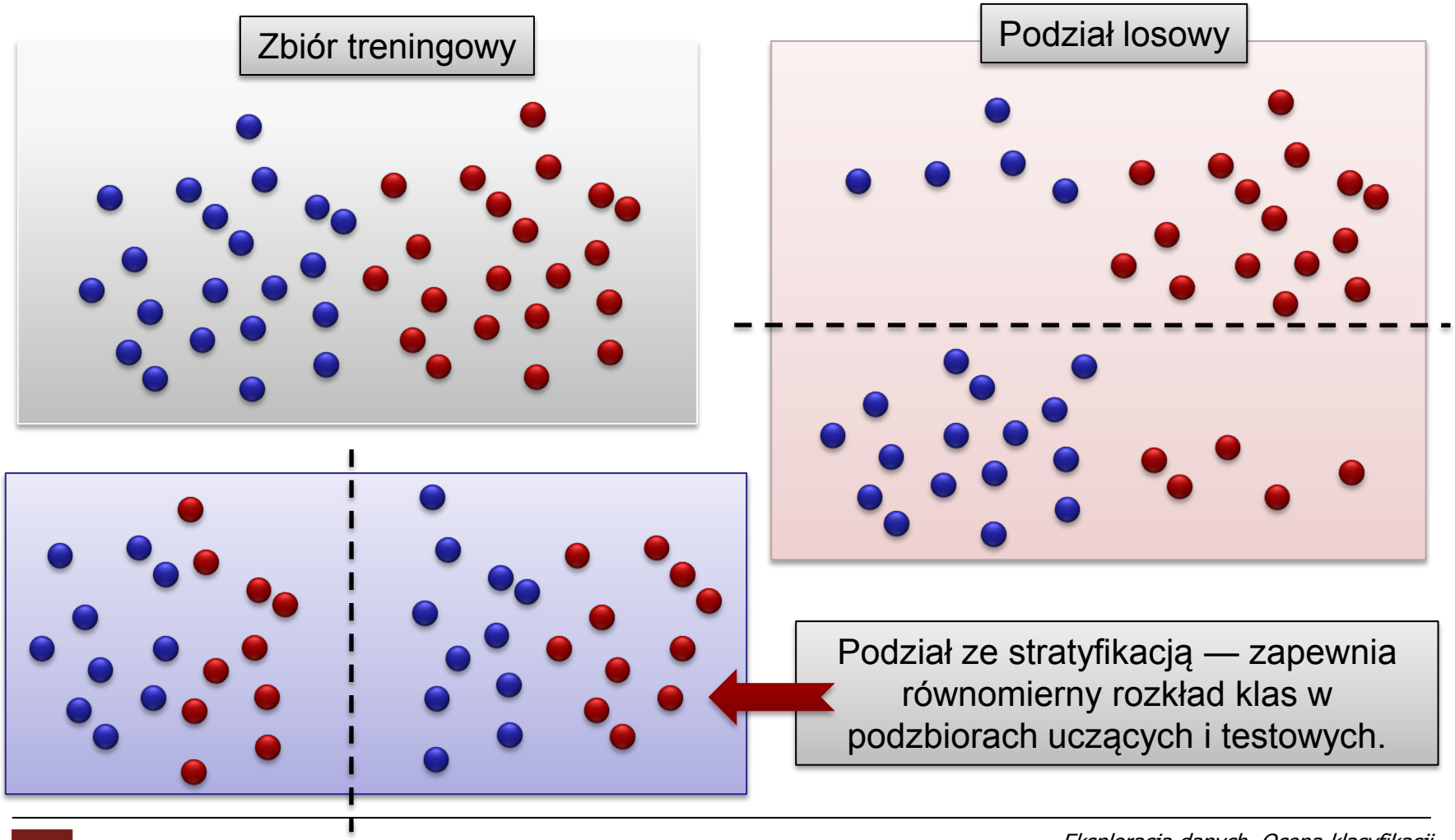
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Próba losowa wektorów danych



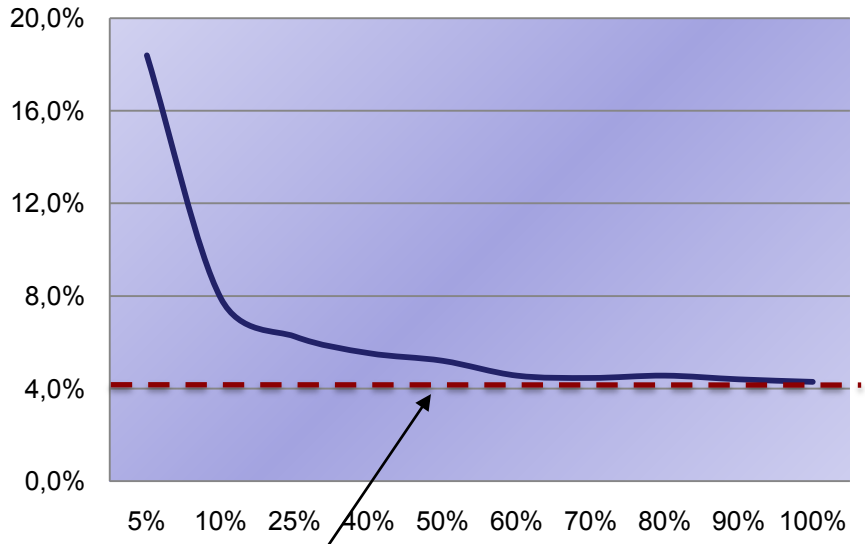
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Walidacja krzyżowa ze stratyfikacją



Krzywa uczenia

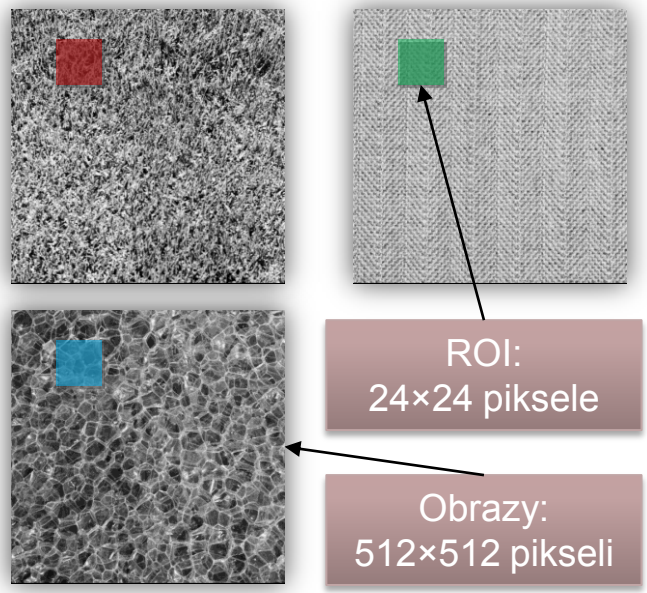
Błąd klasyfikacji w zależności od liczby wektorów



Prawdopodobna wartość błędu rzeczywistego

Procent całkowitej liczby wektorów biorących udział w uczeniu i testowaniu

Przykład



ROI:
24×24 piksele

Obrazy:
512×512 pikseli

Zbiór danych:
- 3 klasy
- 256 wektorów danych w klasie

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

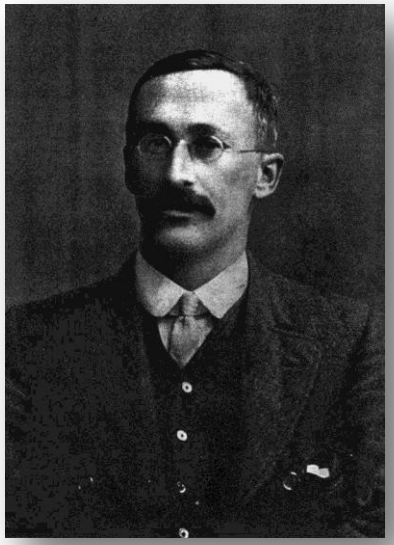
Wybór klasyfikatora

	Zbiory obrazów tekstur (24×24 piksele, 128 wektorów w klasie)		
	grass, weave, plastic	*bark, brick, rafia*	*brick, sand, clothe*
Support vector machine (liniowa funkcja jądra)	**5,9%**	**2,4%**	**2,3%**
Regresja logistyczna	**4,2%**	**1,6%**	**3,9%**
Algorytm 1-R	**21,6%**	**6,3%**	**18,0%**

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Statystyka *t*–Studenta

Hipoteza zerowa: różnica między uzyskanymi wskaźnikami poprawności klasyfikacji jest przypadkowa; w rzeczywistości wszystkie testowane algorytmy są jednakowo dokładne.



Wiliam Seally Gosset, ps. Student, 1876–1937

Statystka *t*–Studenta pozwala zweryfikować słuszność tej hipotezy.

Zbiory danych (jednakowo liczne, z tej samej dziedziny)

	#1	#2	…	#k		
Metoda A:	6%	5%	…	2%	$\bar{e}_A$	μ_A
Metoda B:	7%	8%	…	6%	$\bar{e}_B$	μ_B

Rzeczywiste średnie błędów

Błędy klasyfikacji

Próbkowe wartości średnie błędów

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Sparowany test Studenta

W przypadku gdy rozmiary próbek są duże, ich średnie mają rozkład normalny.

…ale rozmiary nie są duże ☹ → Średnie mają rozkład Studenta.

1. Jeśli poszczególne wyniki dla metod A oraz B uzyskano dla tych samych zbiorów danych, to stosujemy sparowany test Studenta.
2. Analizujemy zmienną:

$$\bar{d} = \bar{e}_A - \bar{e}_B$$

3. Jeżeli hipoteza zerowa jest słuszna, to:

$$\mu_A = \mu_B$$

Standaryzacja zmiennej losowej dla rozkładu Studenta

$$t = \frac{\bar{e}_A - \mu_A}{\sqrt{\sigma_{e_A}^2 / k}}$$

$$t = \frac{\bar{d}}{\sqrt{\sigma_d^2 / k}}$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przedziały ufności dla zmiennej *t*–Studenta

Granice przedziałów ufności

Wartości dla przedziałów jednostronnych

Rozkład Studenta

$\Pr[X \geq z]$	z
0,1%	4,30
0,5%	3,25
1%	2,82
5%	1,83
10%	1,38
20%	0,88

Rozkład Guassa

$\Pr[X \geq z]$	z
0,1%	3,09
0,5%	2,58
1%	2,33
5%	1,65
10%	1,28
20%	0,84

Tabela dla k=10 (9 stopni swobody)

1. Decydujemy się na jakiś poziom ufności → 99%.
2. Obliczamy wartość zmiennej t → t_{exp}.
3. Jeżeli $t > z$ lub $t < -z$ to odrzucamy hipotezę zerową.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Możliwe wyniki predykcji klasy

Przypadek 2 klas

Rzeczywista klasa \ Wynik klasyfikacji	Tak (1)	Nie (0)
Tak (1)	TP	FN
Nie (0)	FP	TN

Oznaczenie danej klasy wektorów etykietą TAK lub NIE zależy od kontekstu.

Wielkość błędu klasyfikacji

$$e = 1 - \frac{TP + TN}{TP + TN + FP + FN}$$

Iloraz *TP*

$$\frac{TP}{TP + FN}$$

Iloraz *FP*

$$\frac{FP}{FP + TN}$$

TP — *true positive*

FP — *false positive*

TN — *true negative*

FN — *false negative*

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Macierz pomyłek

Klasyfikator 1-R

Wynik klasyfikacji

Rzeczywista klasa

	A	B	C	Razem
A	113	0	15	128
B	3	125	0	128
B	6	0	122	128
Razem	122	125	137	384

Wyniki uzyskane dla zbioru obrazów tekstur (*bark, brick, rafia*)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	360	93.75 %
Incorrectly Classified Instances	24	6.25 %
Kappa statistic	**0.9063**	

Na ile tak naprawdę klasyfikator jest skuteczny?

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Statystyka Kappa

Klasyfikator losowy

Rzeczywista klasa	Wynik klasyfikacji: A	B	C	Razem
A	45	58	25	128
B	36	37	55	128
B	67	15	46	128
Razem	148	110	126	384

Liczba poprawnych predykcji klasyfikatora 1-*R*

Liczba poprawnych wskazań klasyfikatora losowego

Liczność zbioru danych

$$\kappa = \frac{L_{1-R} - L_{losowy}}{Q - L_{losowy}}$$

$$L_{1-R} = 360 \quad L_{losowy} = 128$$
$$Q = 384 \quad \kappa = 0{,}9063$$

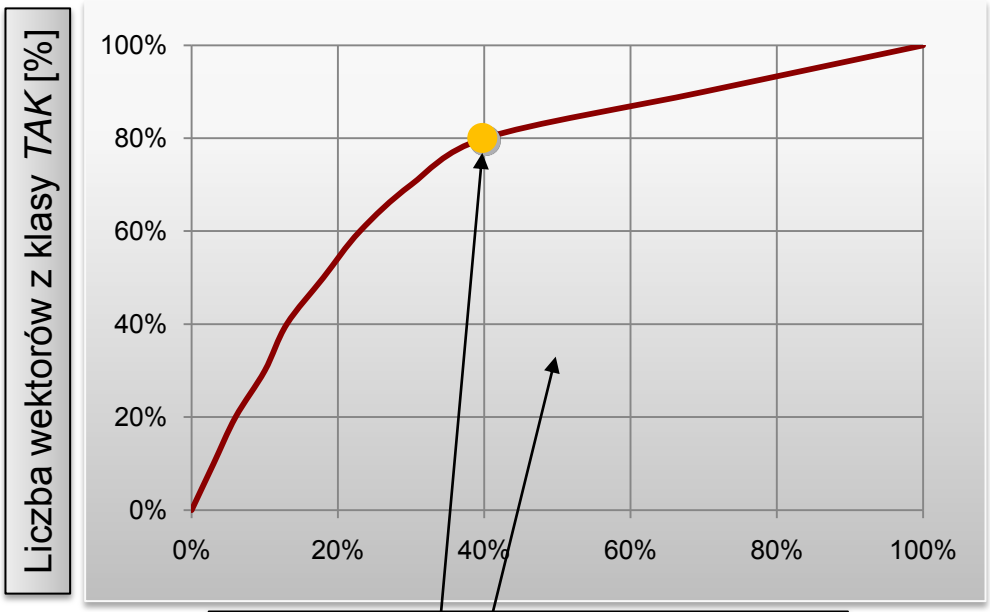
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Wykres wzrostu

Naiwny klasyfikator Bayesa

Pozycja w rankingu	Wynik klasyfikacji	Rzeczywista klasa
1	0,99	TAK
2	0,98	TAK
3	0,97	TAK
4	0,95	NIE
5	0,94	TAK
…	…	…

Wektory uszeregowane zgodnie z malejącym prawdopodobieństwem przynależności do klasy oznaczonej etykietą *TAK*

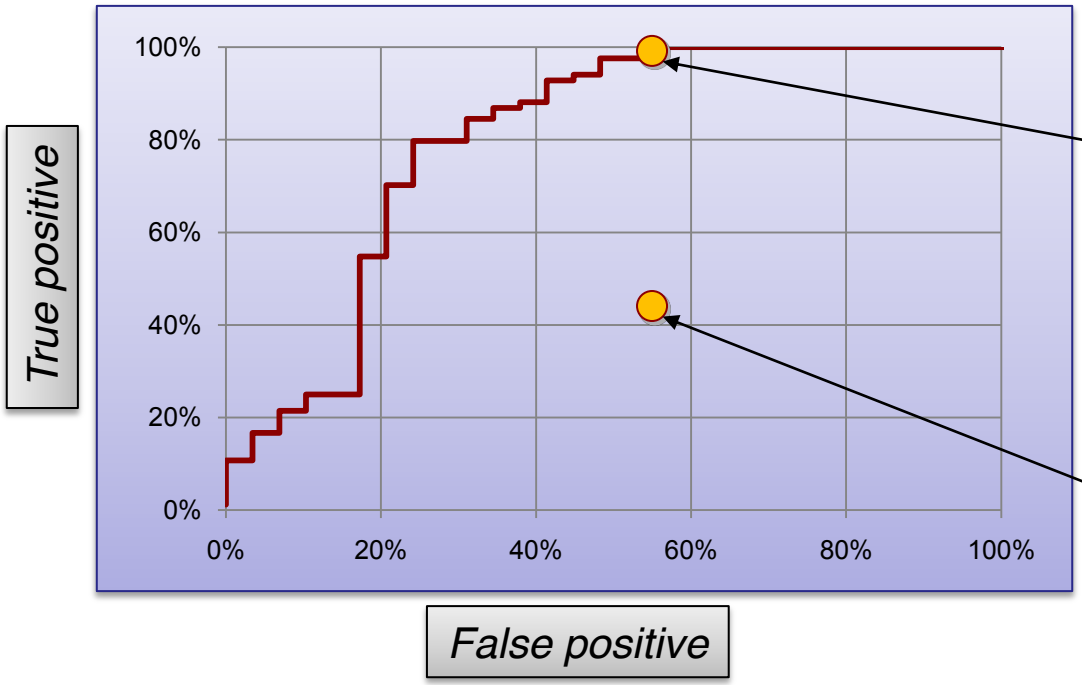


Wystarczy 40% całego zbioru danych, aby znalazło się w nim 80% wszystkich wektorów z klasy „pozytywnej”.

Im większa powierzchnia pod wykresem, tym lepszy klasyfikator.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Krzywa ROC



Znalezienie wszystkich wyników typu *TP* odbywa się kosztem popełnienia ok. 55% błędów typu *FP*.

Podobnie jak w przypadku analizy wykresów wzrostu, powierzchnia pod krzywą ROC dla skutecznego klasyfikatora powinna być jak największa.

ROC (ang. *receiver operating characteristic*) — wielkość używana przy detekcji sygnałów transmitowanych przez zaszumiony kanał ; pozwala określić kompromis pomiędzy liczbą właściwych trafień oraz fałszywych alarmów.