

SZTUCZNA INTELIGENCJA

WYKŁAD 6. SZTUCZNE SIECI NEURONOWE – KLASYFIKACJA

Dr hab. inż. Grzegorz Dudek
Wydział Elektryczny
Politechnika Częstochowska

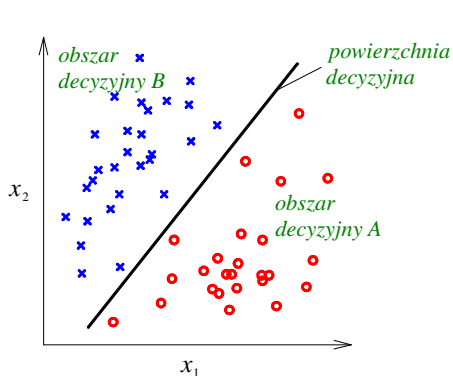
Częstochowa 2014

PROBLEM KLASYFIKACJI DANYCH

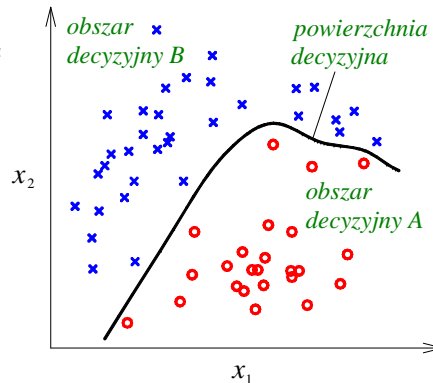
Klasyfikacja – polega na przydzieleniu obserwacji do odpowiedniej klasy. Dostępna jest sklasyfikowana próba ucząca (zbiór uczący), czyli zbiór N przykładów etykietowanych. Przykłady opisane są za pomocą n atrybutów $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$. Na etapie uczenia, wykorzystując informacje zawarte w zbiorze uczącym, określa się charakterystyki klas. Na etapie rozpoznawania dokonuje się klasyfikacji nowych przykładów (zbiór testowy) w oparciu o wyuczone charakterystyki klas.

Typowe zadania klasyfikacji:

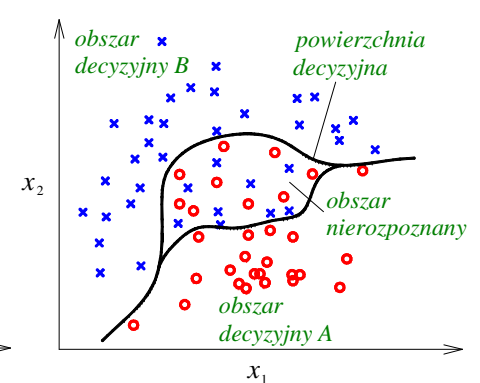
Zbiory separowalne liniowo



Zbiory separowalne nieliniowo

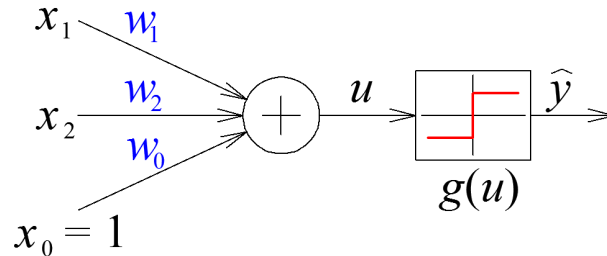


Zbiory nieseparowalne



NEURON JAKO KLASYFIKATOR

Neuron ze skokową funkcją aktywacji pełni funkcję klasyfikatora liniowego



Neuron realizuje funkcję:

$$g(u) = \begin{cases} +1 & \text{jeśli } u = w_1x_1 + w_2x_2 + w_0 \geq 0 \\ -1 & \text{jeśli } u = w_1x_1 + w_2x_2 + w_0 < 0 \end{cases}$$

+1 na wyjściu neuronu oznacza klasę A, -1 oznacza klasę B.

NEURON JAKO KLASYFIKATOR

+1 pojawi się, gdy $w_1x_1 + w_2x_2 + w_0 \geq 0$, tzn.: $x_2 \geq -\frac{w_1}{w_2}x_1 - \frac{w_0}{w_2}$

-1 pojawi się, gdy: $x_2 < -\frac{w_1}{w_2}x_1 - \frac{w_0}{w_2}$

Powyższe nierówności definiują półpłaszczyzny – obszary decyzyjne obu klas. Linia decyzyjna rozdzielająca te obszary ma postać:

$$x_2 = -\frac{w_1}{w_2}x_1 - \frac{w_0}{w_2}$$

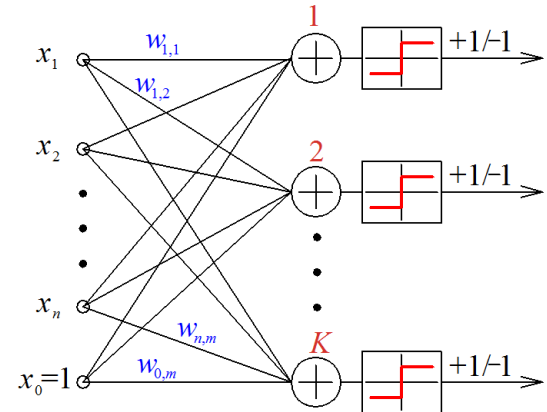
Jak widać wagi sieci określają współczynnik kierunkowy i wyraz wolny linii decyzyjnej. W procesie uczenia wagi zostaną tak dobrane, aby prosta separowała obie klasy (jeśli to możliwe).

NEURONOWY KLASYFIKATOR WIELOKLASOWY

Gdy liczba klas jest większa od 2, wynosi K , stosujemy K neuronów. Każdy neuron reprezentuje inną klasę. Klasa i -ta sygnalizowana jest wartością $+1$ na wyjściu i -tego neuronu. Pozostałe neurony na wyjściach mają wartość -1 .

W tym przypadku etykieta klasy ma postać wektora o K składowych: $\mathbf{y} = [y_1, y_2, \dots, y_K]$, $y_i = \pm 1$.

Symbole klas można zakodować na mniejszej liczbie bitów, np. kl. 1 – 00, kl. 2 – 01, kl. 3 – 10, kl. 4 – 11.



Wagi sieci adaptuje się w procesie uczenia według wzoru:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i + \frac{1}{2} \eta (y_i - \hat{y}_i) \mathbf{x},$$

gdzie i to numer neuronu (klasy).

ALGORYTM UCZENIA KLASYFIKATORA DYSKRETNEGO DLA WIELU KLAS

Dane jest p przykładów uczących: $\{\langle \mathbf{x}_1, \mathbf{y}_1 \rangle, \langle \mathbf{x}_2, \mathbf{y}_2 \rangle, \dots, \langle \mathbf{x}_p, \mathbf{y}_p \rangle\}$

1. Wybór $\eta > 0$ i wag startowych.
2. Podanie przykładu na wejście i obliczenie sygnału wyjściowego:

$$\hat{y}_i = \text{sgn}(\mathbf{w}_i^T \mathbf{x}), \quad i=1,2,\dots,K,$$

gdzie $\text{sgn}(u)=+1$, gdy $u > 0$; 0 , gdy $u = 0$ lub -1 , gdy $u < 0$.

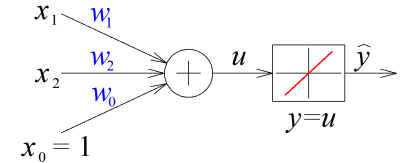
3. Obliczenie nowych wag:

$$\mathbf{w}_i \leftarrow \mathbf{w}_i + \frac{1}{2} \eta (y_i - \hat{y}_i) \mathbf{x},$$

4. Jeśli warunek zatrzymania jest niespełniony przejście do kroku 2.

MASZYNA LINIOWA

Maszyna liniowa klasyfikuje dane liniowo separowalne należące do K klas. Neurony mają liniowe funkcje aktywacji. Każdy neuron reprezentuje inną klasę.



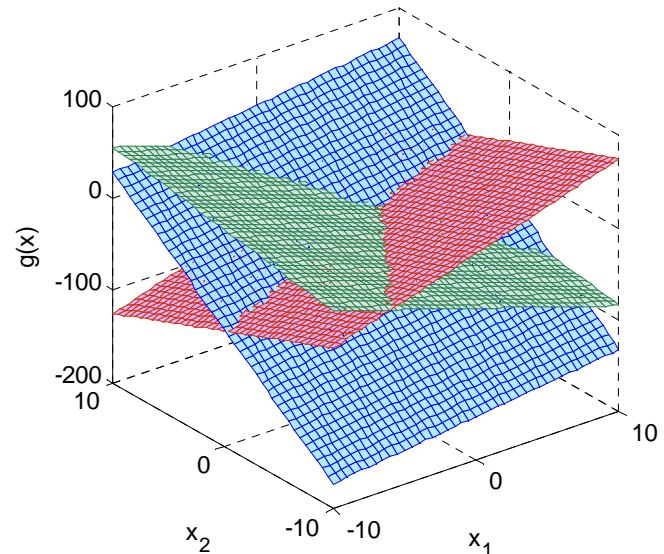
Funkcja realizowana przez neuron (**funkcja dyskryminacyjna**):

$$g(\mathbf{x}) = w_1 x_1 + w_2 x_2 + w_0$$

W przypadku trzech klas otrzymujemy trzy funkcje dyskryminacyjne realizowane przez neurony.

Należy tak dobrać wagi, aby dla przykładów \mathbf{x} z klasy i -tej wartość i -tej funkcji była większa od wartości funkcji reprezentujących pozostałe klasy:

$$\mathbf{x} \in c_i \Rightarrow g_i(\mathbf{x}) = \max_l g_l(\mathbf{x}).$$



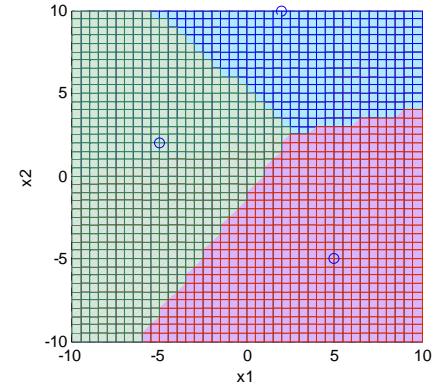
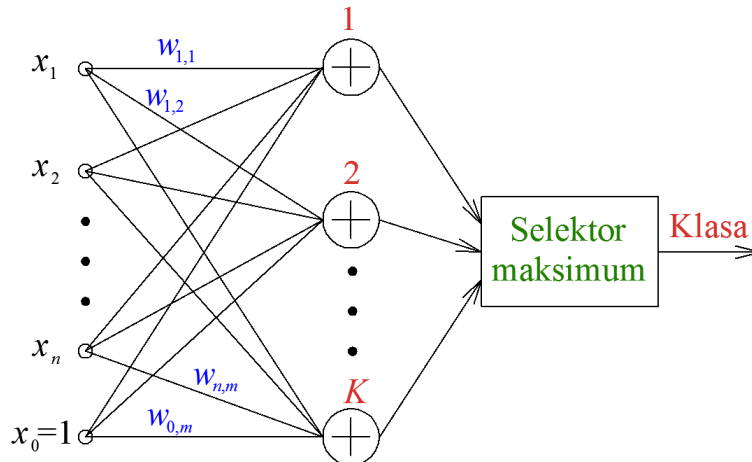
MASZYNA LINIOWA

Powierzchnie decyzyjne separujące dwie klasy powstają w miejscu przecięcia funkcji dyskryminacyjnych dla tych klas:

$$g_i(\mathbf{x}) = g_j(\mathbf{x}).$$

Wagi można dobrać w sposób analityczny na podstawie punktów prototypowych reprezentujących poszczególne klasy [Żur].

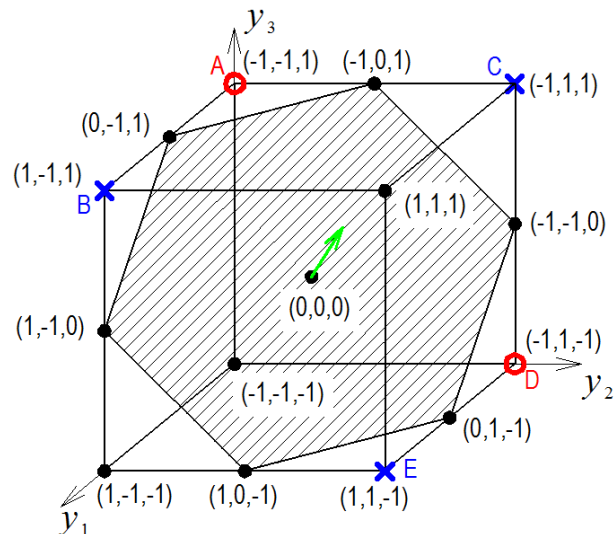
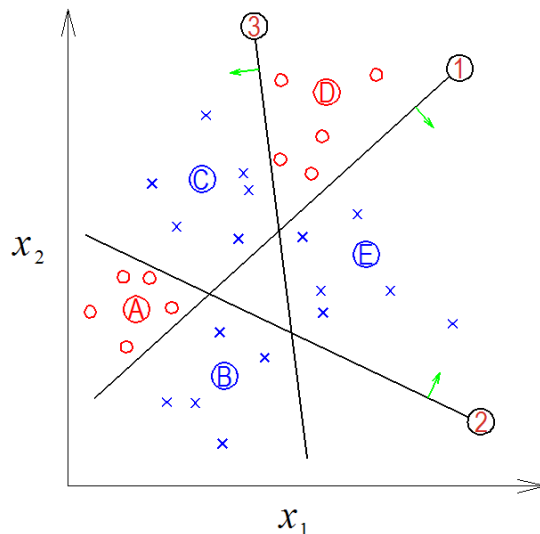
Schemat maszyny liniowej:



KLASYFIKACJA OBRAZÓW LINIOWO NIESEPAROWALNYCH

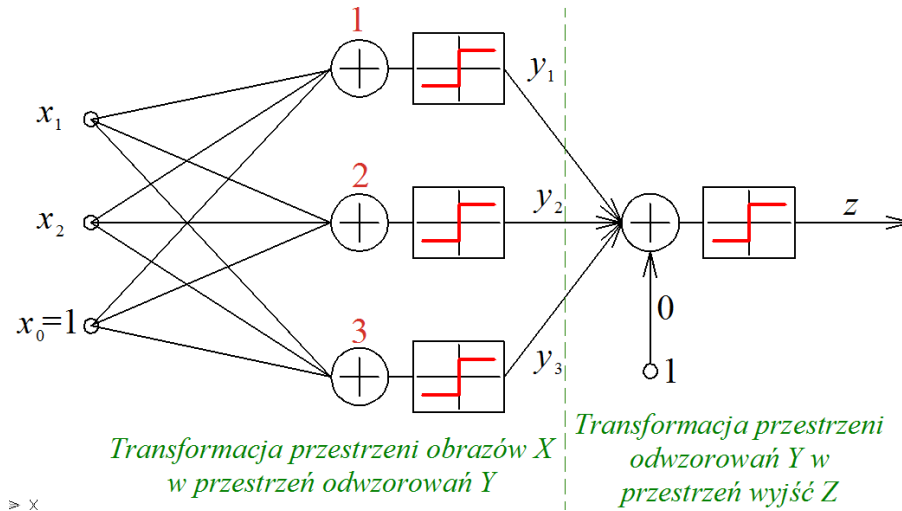
Każdy neuron ze skokową funkcją aktywacji dzieli liniowo płaszczyznę na dwie części, tak aby wydzielone obszary decyzyjne zawierały przykłady z jednej klasy.

Na wyjściu m neuronów otrzymujemy wektor $\mathbf{y} = [y_1, y_2, \dots, y_m]$, $y_i = \pm 1$. Zachodzi transformacja n -wymiarowych przykładów \mathbf{x} (**przestrzeń obrazów**) w m -wymiarową przestrzeń wektorów wyjściowych \mathbf{y} (**przestrzeń odwzorowań**). Przykłady w tej nowej przestrzeni są separowalne za pomocą płaszczyzny realizowanej przez neuron drugiej warstwy.



KLASYFIKACJA OBRAZÓW LINIOWO NIESEPAROWALNYCH

Model sieci dwuwarstwowej do klasyfikacji obrazów liniowo nieseparowalnych.

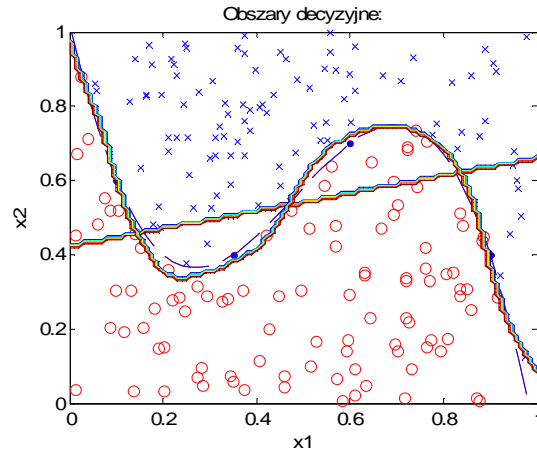


NIELINIOWE POWIERZCHNIE DECYZYJNE

Do tworzenia bardziej złożonych, nieliniowych powierzchni decyzyjnych stosuje się sieci wielowarstwowe z nieliniowymi funkcjami aktywacji.

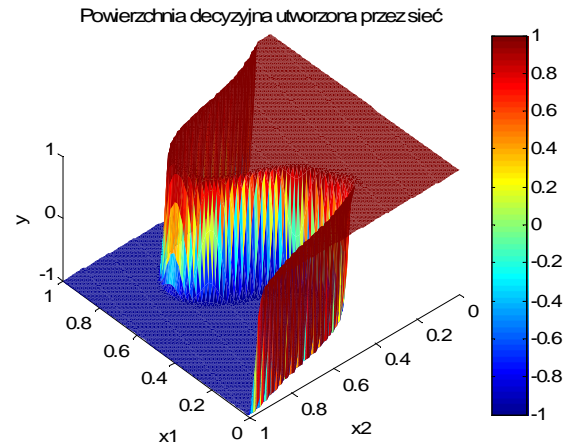
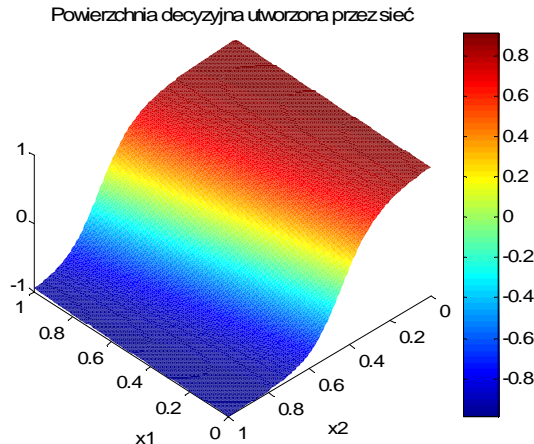
Im więcej neuronów tym powierzchnie decyzyjne mogą być bardziej złożone.

Linie decyzyjne utworzone przez sieci z jednym (wariant 1) i pięcioma (wariant 2) neuronami w warstwie ukrytej:



NIELINIOWE POWIERZCHNIE DECYZYJNE

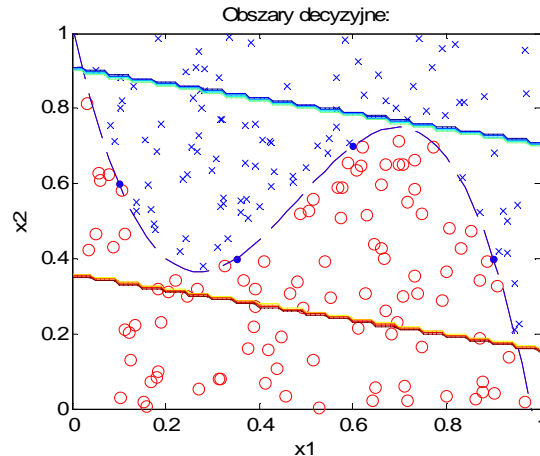
W trakcie treningu minimalizowany jest błąd średniokwadratowy pomiędzy pożądanym numerem klasy (+1 lub -1), a odpowiedzią sieci. Przy sigmoidalnej bipolarnej funkcji aktywacji neuronu wyjściowego odpowiedź sieci jest liczbą rzeczywistą z zakresu od -1 do +1. Dla zamieszczonych powyżej danych uczących powierzchnie odpowiedzi sieci z jednym i pięcioma neuronami wyglądają następująco:



NIELINIOWE POWIERZCHNIE DECYZYJNE

Linie decyzyjne powstają z przekroju powierzchni odpowiedzi sieci płaszczyzną $y = 0$, co oznacza, że jeśli sieć daje odpowiedź dodatnią przyjmuje się klasę A, a jeśli ujemną – klasę B.

Można przyjąć inną zasadę – jeśli odpowiedź sieci jest powyżej $+0,8$ oznacza to klasę A, a jeśli poniżej $-0,8$ – oznacza to klasę B. Odpowiedzi w przedziale $[-0,8; +0,8]$ uznawane są jako brak decyzji (obszar nierozpoznany). W takim przypadku linie decyzyjne utworzone przez sieć z jednym neuronem (wariant 3) mogą wyglądać tak:



OCENA KLASYFIKATORA

Wyniki klasyfikacji danych w opisywanych powyżej wariantach sieci zamieszczono w tabeli:

Wariant sieci	Odsetek poprawnych klasyfikacji	Odsetek błędnych klasyfikacji	Odsetek nierozstrzygniętych klasyfikacji
1	82.50%	17.50%	0.00%
2	98.00%	2.00%	0.00%
3	47.50%	1.00%	51.50%

Macierz przekłamań:

Klasa przypisana	0	TP	FP	PPV
	1	FN	TN	NPV
		TPR	SPC	ACC
	Klasa prawdziwa	0	1	

TP (true positive) – liczba przypadków z klasy 0 poprawnie zaklasyfikowanych,

TN (true negative) – liczba przypadków z klasy 1 poprawnie zaklasyfikowanych,

FP (false positive) – liczba przypadków z klasy 0 błędnie zaklasyfikowanych,

FN (false negative) – liczba przypadków z klasy 1 błędnie zaklasyfikowanych,

Czułość

$$TPR = TP / (TP + FN)$$

Specyficzność

$$SPC = TN / (FP + TN)$$

Precyzja

$$PPV = TP / (TP + FP)$$

Negative predictive value

$$NPV = TN / (TN + FN)$$

Dokładność

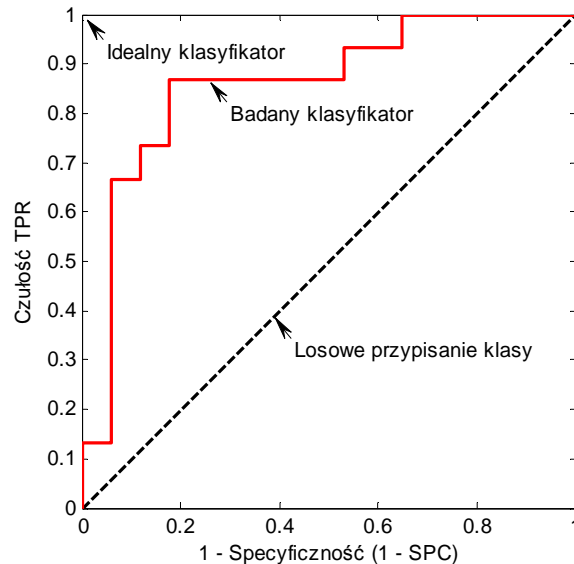
$$ACC = (TP + TN) / (TP + TN + FP + FN)$$

OCENA KLASYFIKATORA

Krzywa ROC

Training Confusion Matrix			Validation Confusion Matrix		
Output Class	Target Class		Output Class	Target Class	
	0	1		0	1
0	65 42.8%	1 0.7%	0	8 25.0%	0 0.0%
1	3 2.0%	83 54.6%	1	2 6.3%	22 68.8%
	95.6% 4.4%	98.8% 1.2%		80.0% 20.0%	100% 0.0%
		97.4% 2.6%			93.8% 6.3%

Test Confusion Matrix			All Confusion Matrix		
Output Class	Target Class		Output Class	Target Class	
	0	1		0	1
0	14 43.8%	3 9.4%	0	87 40.3%	4 1.9%
1	3 9.4%	12 37.5%	1	8 3.7%	117 54.2%
	82.4% 17.6%	80.0% 20.0%		91.6% 8.4%	96.7% 3.3%
		81.3% 18.8%			94.4% 5.6%



Macierze przekłamań można tworzyć także dla wielu klas.