

Wydział Zarządzania AGH

Katedra Informatyki Stosowanej

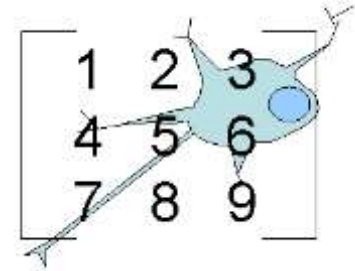


Sztuczne sieci neuronowe

Systemy wspomagania decyzji

Sztuczne sieci neuronowe

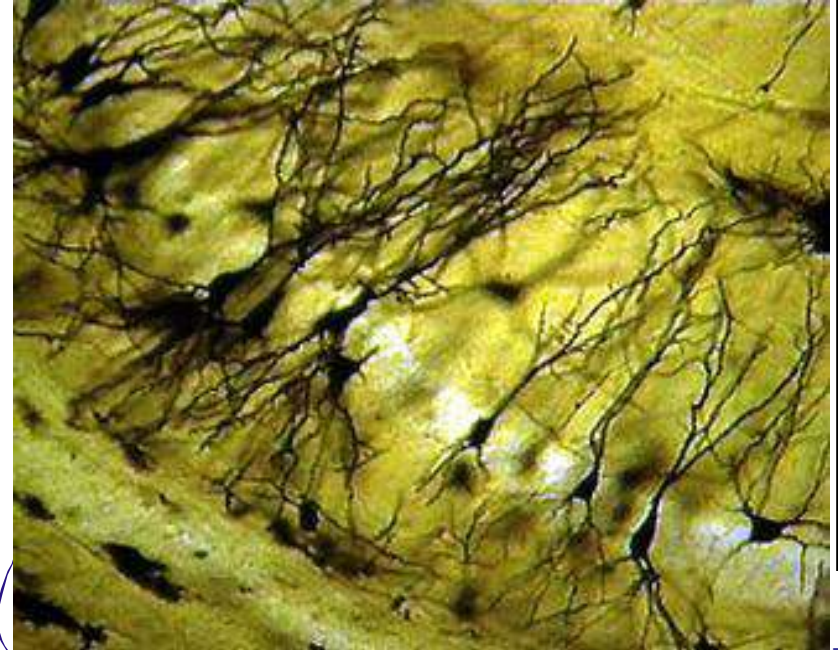
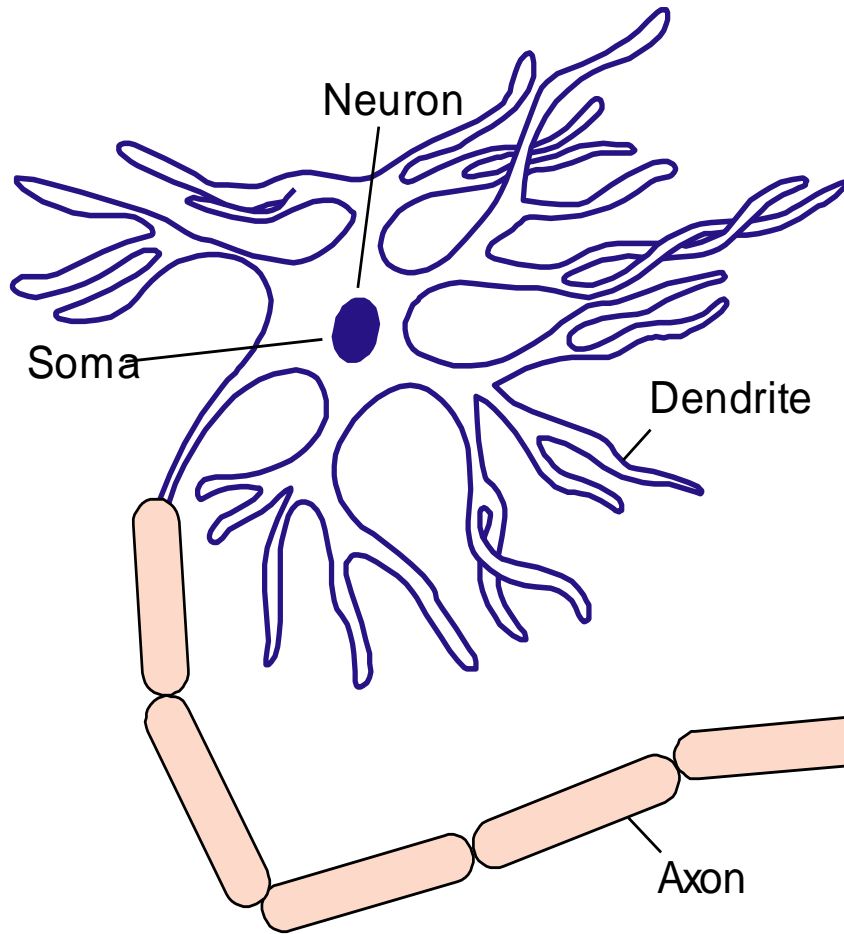
- Wprowadzenie
- Trochę historii
- Podstawy działania
- Funkcja aktywacji
- Uczenie sieci
- Typy sieci
- Zastosowania



Wprowadzenie

- Zainteresowanie SSN wynika z potrzeby budowania efektywnych systemów przetwarzania informacji opartych na regułach neurofizjologii. Przetwarzanie w takich systemach odbywa się w sposób równoległy w odróżnieniu od klasycznych, sekwencyjnych systemów wywodzących się z modelu von Neumanna. Założenia:
 - tradycyjne systemy są mało wydajne,
 - mózg ludzki ma przewagę nad maszyną.
- Często podkreślane podobieństwo do systemów biologicznych jest raczej nieuzasadnione, gdyż wciąż zbyt mało wiemy o rzeczywistym funkcjonowaniu mózgu, a sztuczne sieci są z pewnością znacznie uproszczonym modelem systemu neuronów.

Wprowadzenie



Sztuczne sieci neuronowe

Wprowadzenie

- Liczba neuronów: $\sim 8,6 \cdot 10^{10}$
- Liczba połączeń poj. neuronu: $\sim 10^4 - 10^5$
- Czas reakcji: $\sim 0,001$ sekundy
- Liczba operacji neuronu: $10-100/s$
- Szybkość pracy mózgu $\sim 10^{15} - 10^{17}$ op/s

- W rezultacie:
czas rozpoznania obrazu: $\sim 0,1$ sekundy

Wprowadzenie - trochę historii

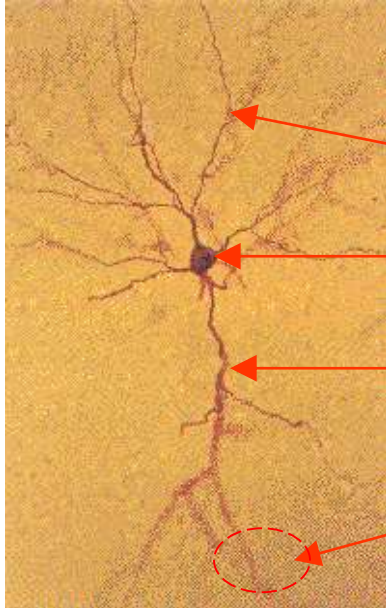
- w 1943 McCulloch i Pitts opracowali matematyczny model sztucznego neuronu.
- w 1957 Rosenblatt zbudował pierwszy (hardwarowy) model neuronu przeznaczony do rozpoznawania znaków alfanumerycznych.
- w 1969 Minsky i Papert udowodnili, że perceptron ma bardzo ograniczone zastosowanie.
- Lata 80. to model sieci wielowarstwowej i efektywny algorytm uczenia sieci.
- Lata 90. - eksplozja. Ponad 80% firm z listy Fortune 500 ma programy R&D z dziedziny SSN:
 - tysiące publikacji,
 - komercyjne pakiety aplikacyjne.

Podstawy działania

- Cechy charakterystyczne:
 - zbiór elementów przetwarzających,
 - stan aktywacji a_i (wejście) dla każdego elementu,
 - połączenia między elementami; ogólnie rzecz biorąc każde połączenie jest zdefiniowane przez wagi w_{ij} , które określają efekt sygnału (wpływ) elementu j na element i ,
 - reguła propagacji, która określa wejście i_i elementu na podstawie sygnałów wejściowych,
 - funkcja aktywacji F_i określająca nowy poziom wyjścia elementu na podstawie wejścia $i_i(t)$,
 - zewnętrzne wejście z_i dla każdego elementu i .
- Poniższe rysunki ilustrują te zasady.

Podstawy działania

NEURON



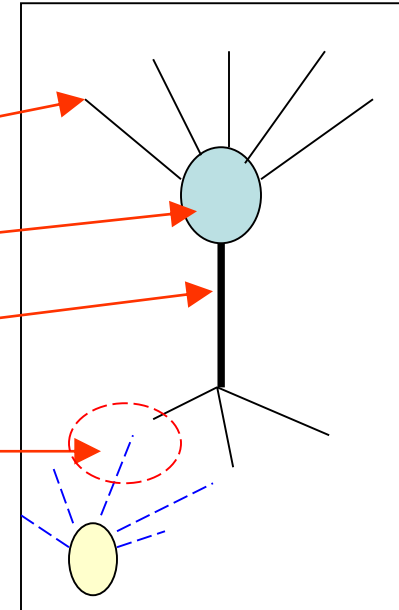
Wygląd

Dendryt

Ciało

Akson

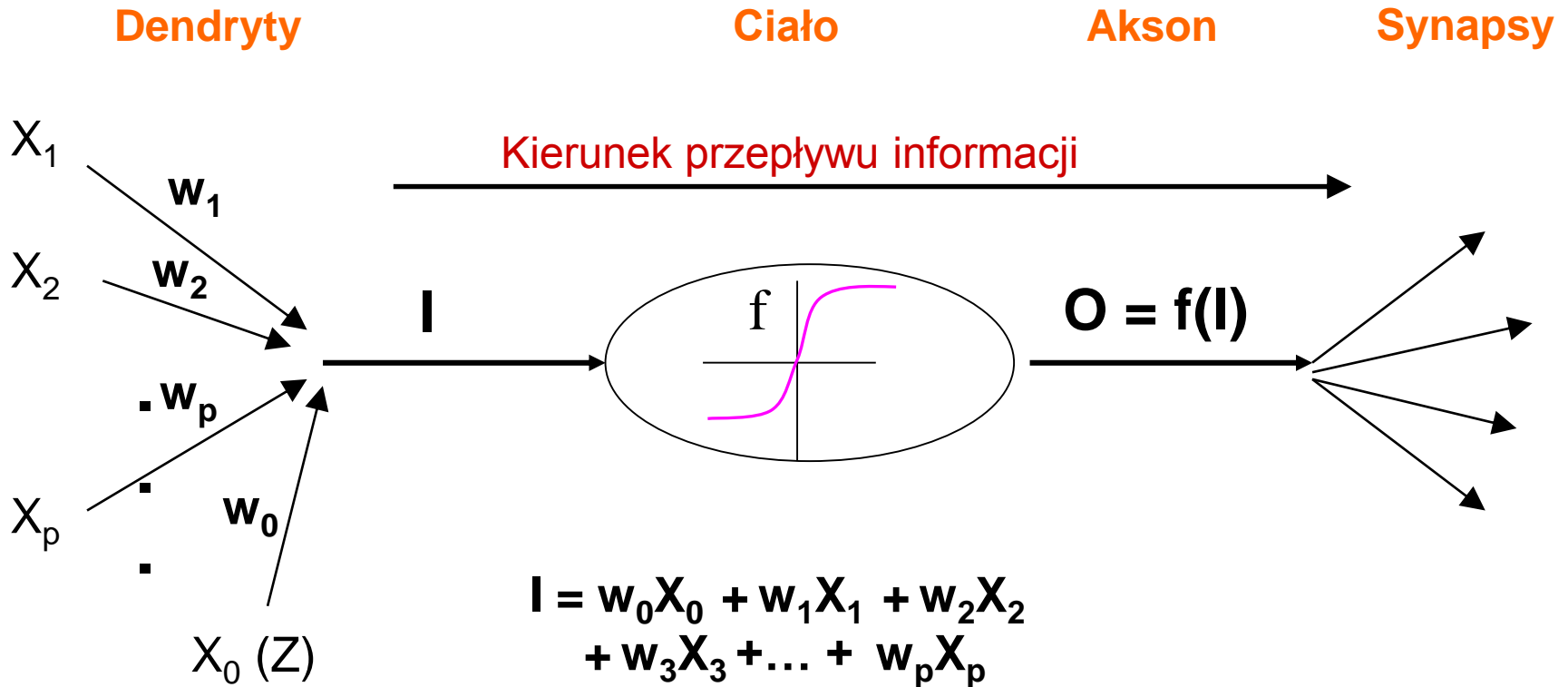
Synapsa



Schemat

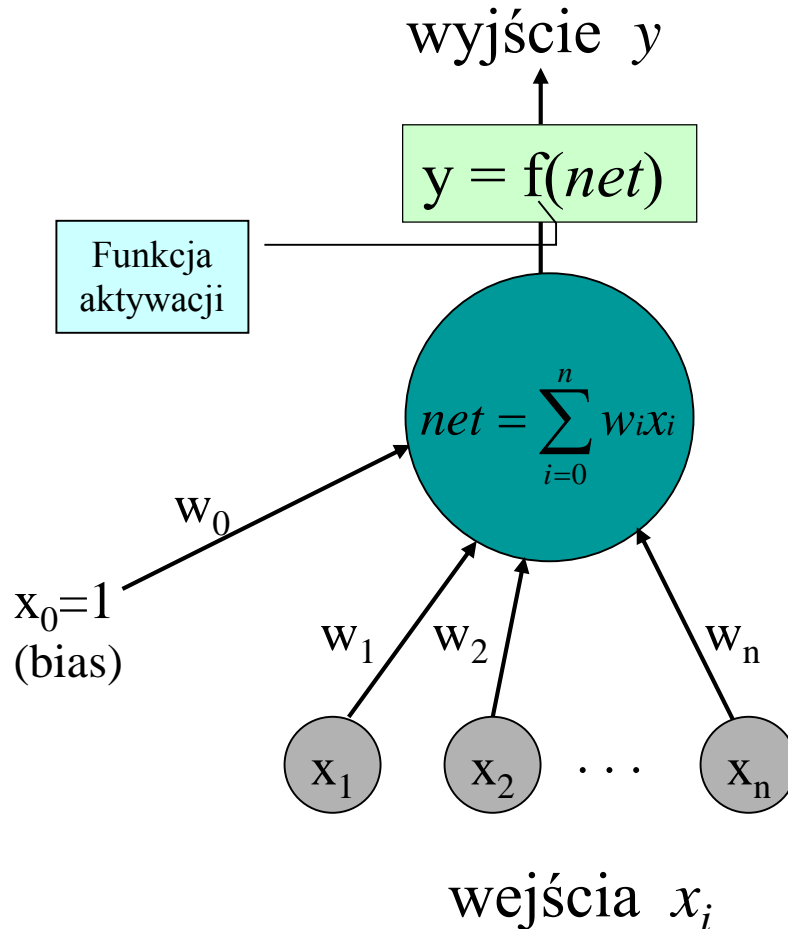
- **Dendryty** – otrzymują informacje
- **Ciało** – przetwarza je
- **Akson** – dostarcza przetworzone informacje do innych neuronów
- **Synapsa** – połączenie między synapsami i dendrytami różnych neuronów

Podstawy działania



Podstawy działania

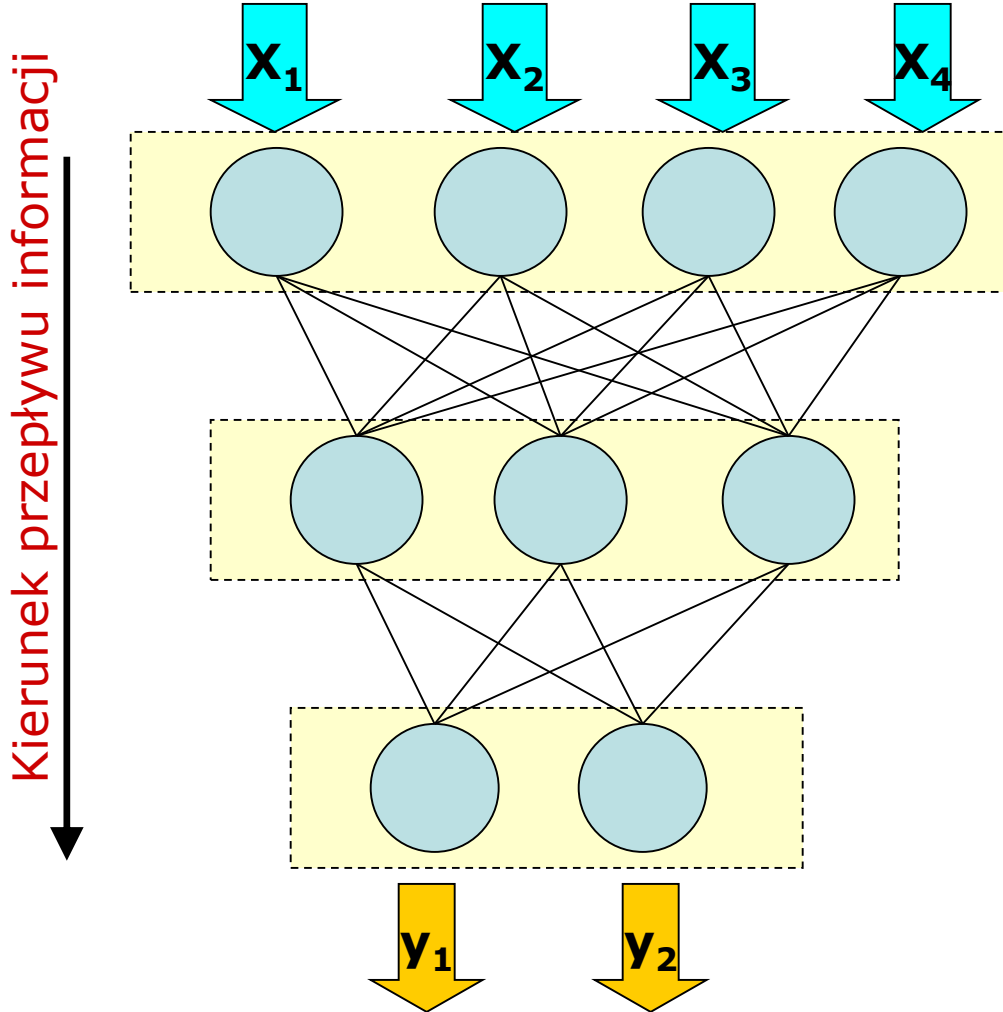
Element przetwarzający - perceptron



Zadaniem elementu jest przetworzyć otrzymany sygnał (od sąsiadów lub z zewnątrz) na sygnał wyjściowy i rozesłać go do innych elementów.

Podstawy działania

Zbiór neuronów tworzy tzw. warstwę



Warstwa wejściowa

każdy neuron otrzymuje **tylko** jedno wejście, bezpośrednio z zewnątrz

Warstwa ukryta

łączy warstwę wejściową z wyjściową

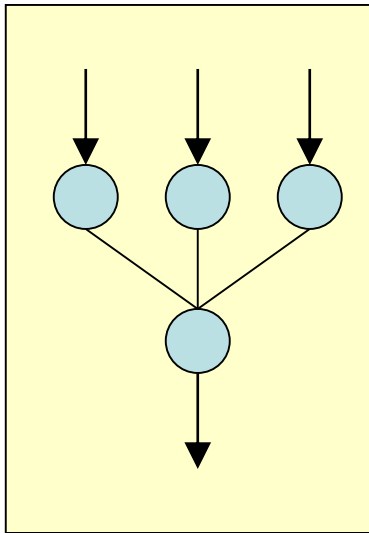
Warstwa wyjściowa

wyjście każdego neuronu skierowane bezpośrednio na zewnątrz

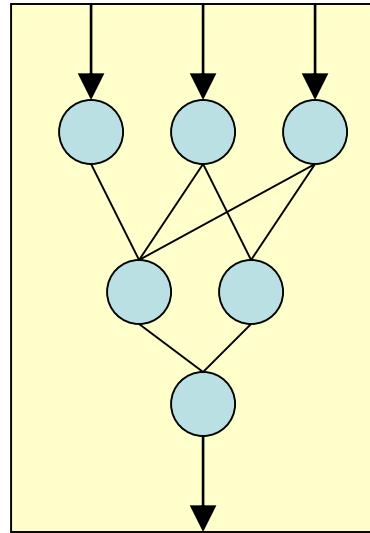
Podstawy działania

Liczba ukrytych warstw może być różna

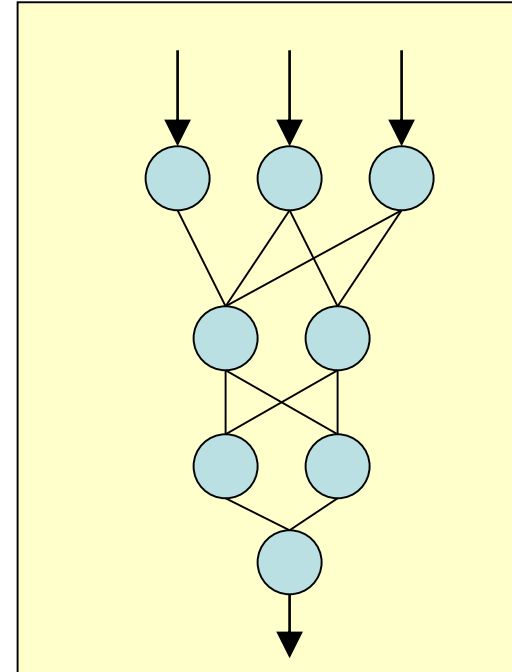
bez



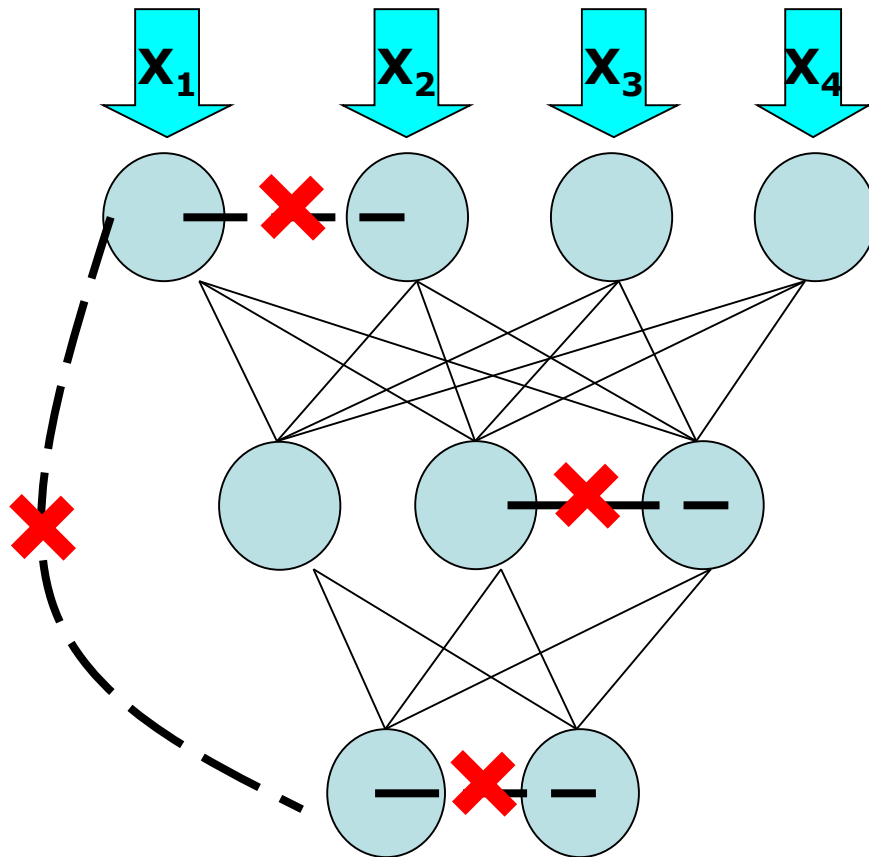
jedna



wiele



Podstawy działania



Uwaga

wewnątrz warstwy neurony **nie** są ze sobą połączone

neurony z danej warstwy są połączone **tylko** z neuronami warstwy **następnej** (feed-forward)

omijanie warstwy **nie** jest dozwolone

Podstawy działania

Połączenia elementów

Przyjmuje się, że każdy element daje addytywny wkład do elementu, z którym jest połączony:

$$i_i(t) = \sum_j w_{ij}(t) \bullet x_j(t) + z_i(t)$$

Jeśli waga w_{ij} jest dodatnia mówimy o pobudzeniu elementu i ze strony elementu j , jeśli ujemna - mówimy o powstrzymywaniu.

Podstawy działania

Aktywacja i obliczanie wyjścia

Funkcja F daje nową wartość aktywacji elementu i , na podstawie wejścia $i_i(t)$

$$o_i(t + 1) = F_i(i_i(t))$$

W zależności od charakterystyki funkcji aktywacji na wyjściu elementu i otrzymujemy różne wartości.

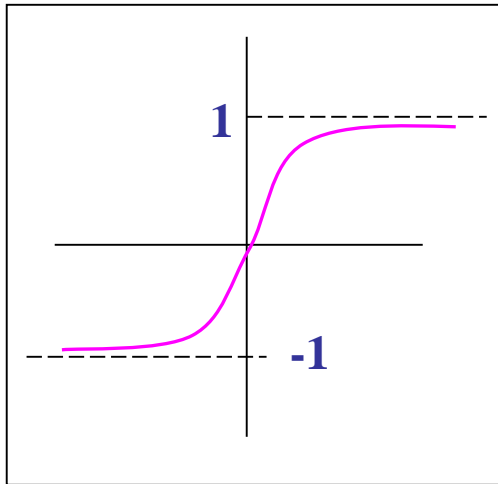
Podstawy działania

Aktywacja i obliczanie wyjścia

- Wybór funkcji aktywacji zależy od rodzaju problemu, jaki stawiamy do rozwiązania przed siecią. Dla sieci wielowarstwowych najczęściej stosowane są funkcje nieliniowe, gdyż neurony o takich charakterystykach wykazują największe zdolności do nauki oraz dają możliwość odwzorowania dowolnej zależności pomiędzy wejściem a wyjściem sieci. Umożliwia to otrzymanie na wyjściu sieci informacji ciągłej, a nie tylko postaci: TAK - NIE.
- Wymagane cechy funkcji aktywacji to:
 - funkcja ciągła w zadanym przedziale,
 - łatwa do obliczenia i ciągła pochodna,
 - możliwość wprowadzenia do argumentu parametru do ustalania kształtu krzywej.

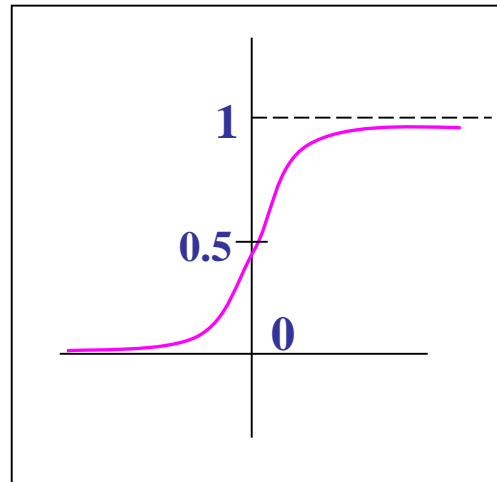
Podstawy działania

Aktywacja i obliczanie wyjścia



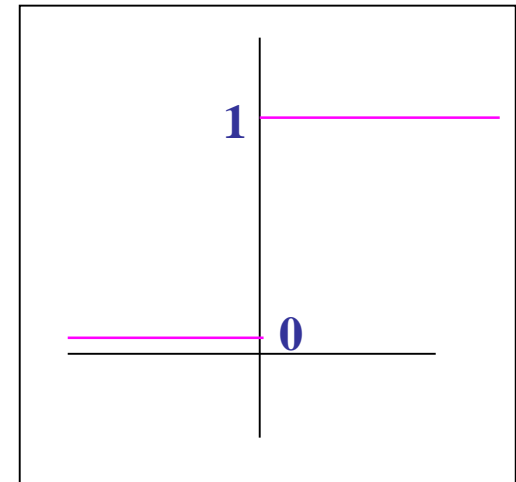
Tanh

$$f(x) = (e^x - e^{-x}) / (e^x + e^{-x})$$



Logistyczna

$$f(x) = 1 / (1 + e^{-x})$$

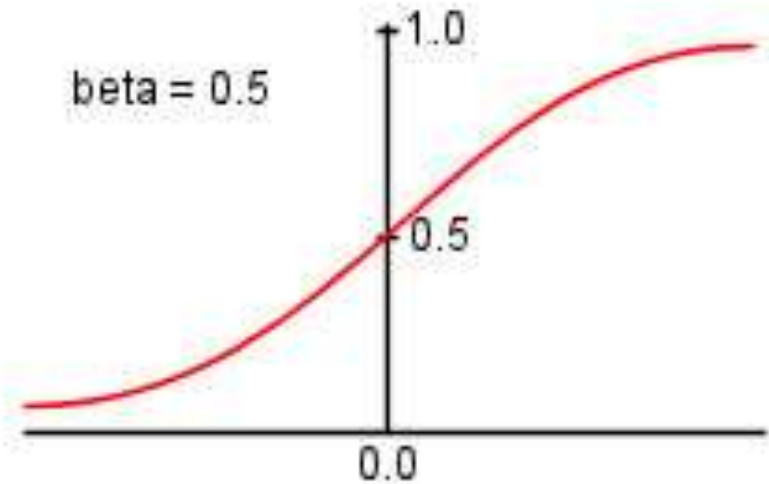
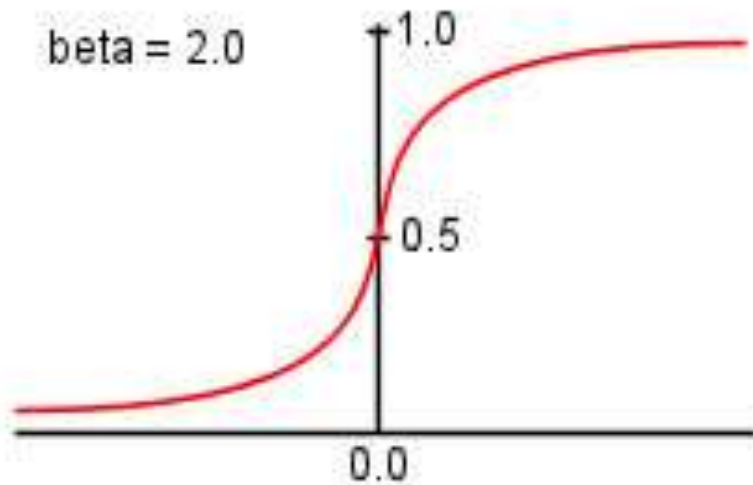


Skok jedn.

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Podstawy działania

Funkcja logistyczna



$$y = f(x) = \frac{1.0}{1.0 + \exp(-(\text{beta} * x))}$$

$$f'(y) = y * (1.0 - y)$$

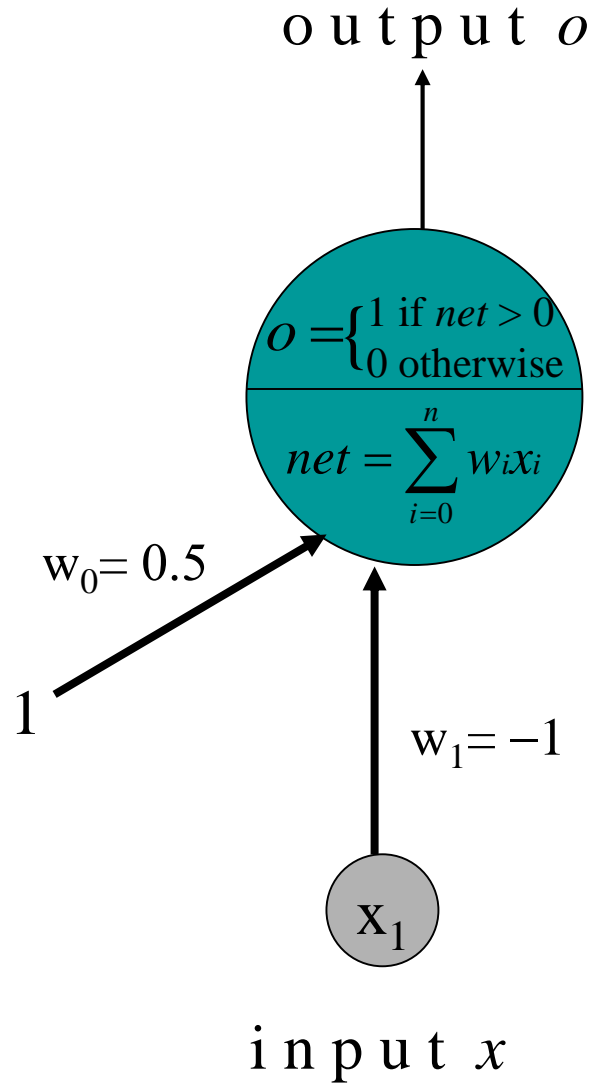
Sieć perceptronowa

- Perceptron zaproponowany przez Rosenblatta w 1959 roku jest siecią jednokierunkową złożoną z elementów binarnych realizowaną wg następujących zasad:
 - elementem składowym jest sztuczny neuron opisany funkcją aktywacji *skok jednostkowy*.
 - sieć można podzielić na uporządkowane i rozłączne podzbiory elementów, zwane warstwami: wejściową i wyjściową.
 - perceptron nie zawiera połączeń między elementami tej samej warstwy.
 - połączenia między warstwami skierowane są zgodnie z ich uporządkowaniem.

Sieć perceptronowa

Perceptron
odwracanie

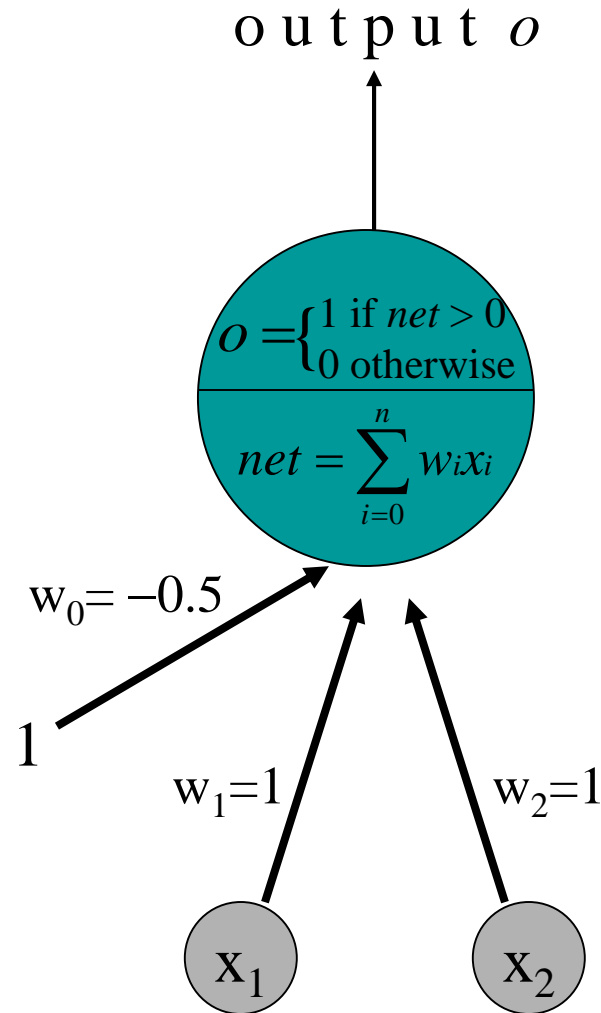
input x_1	output
0	1
1	0



Sieć perceptronowa

Perceptron
OR

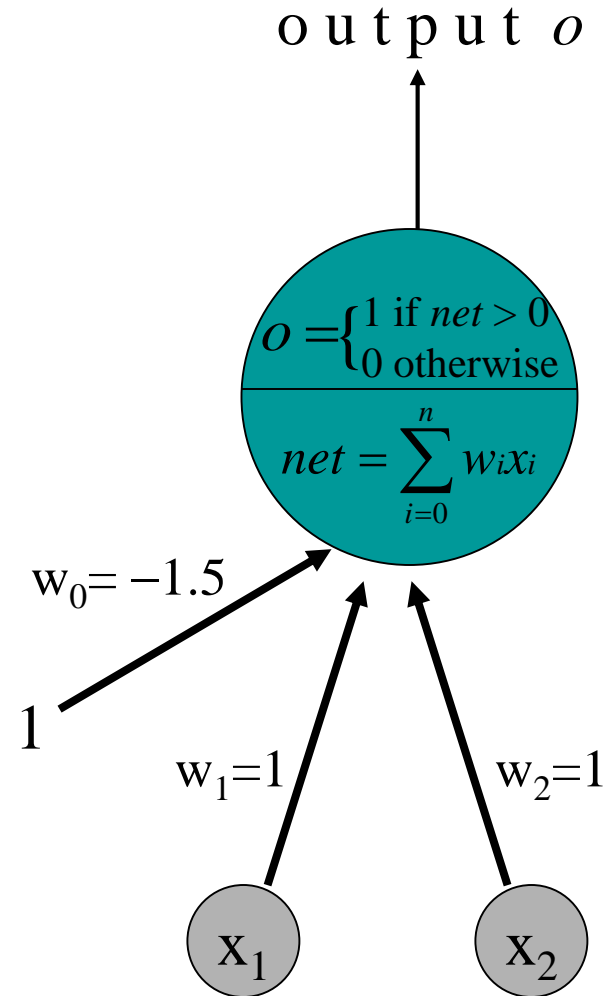
input x1	input x2	output
0	0	0
0	1	1
1	0	1
1	1	1



Sieć perceptronowa

Perceptron
AND

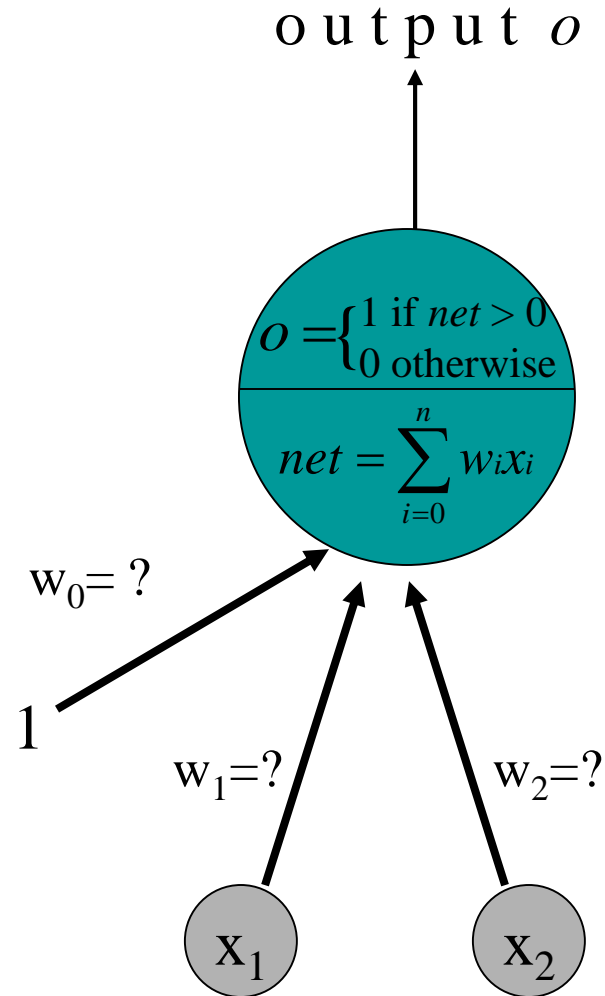
input x1	input x2	output
0	0	0
0	1	0
1	0	0
1	1	1



Sieć perceptronowa

Perceptron XOR

input x1	input x2	output
0	0	0
0	1	1
1	0	1
1	1	0



Metody uczenia

- Uczenie polega na takim wyznaczeniu wartości wag, by sieć rozwiązywała wyuczone zadanie w sposób automatyczny.

Wyróżniamy dwie kategorie:

- **uczenie nadzorowane** (supervised), gdzie dostarcza się wzorce wejściowe i odpowiadające im wyjścia.
- **uczenie nienadzorowane**, gdzie sieć samodzielnie próbuje odkryć statystycznie istotne cechy wzorca; w tej metodzie nie ma danego a priori zbioru kategorii, według których mają być klasyfikowane wzorce.

Metody uczenia

- Wszystkie metody uczenia są wariantami reguły uczenia Hebba opisanej w *Organization of Behaviour* (1949). Podstawową ideą jest tu założenie, że jeśli dwa elementy działają zgodnie, ich połączenie musi być wzmocnione. Jeżeli element i otrzymuje wejście od j , to najprostsza wersja reguły Hebba opisuje zmianę wag w_{ij} o:

$$\Delta w_{ij} = \alpha \bullet o_i \bullet o_j$$

gdzie α jest dodatnią stałą nazywaną *stałą uczenia*.

Metody uczenia

- Inna powszechnie stosowana reguła używa nie aktualnej aktywacji elementu i , lecz różnicy między aktualną a wagami zmieniane są tym silniej, im większy jest błąd pożądaną aktywacją elementu.

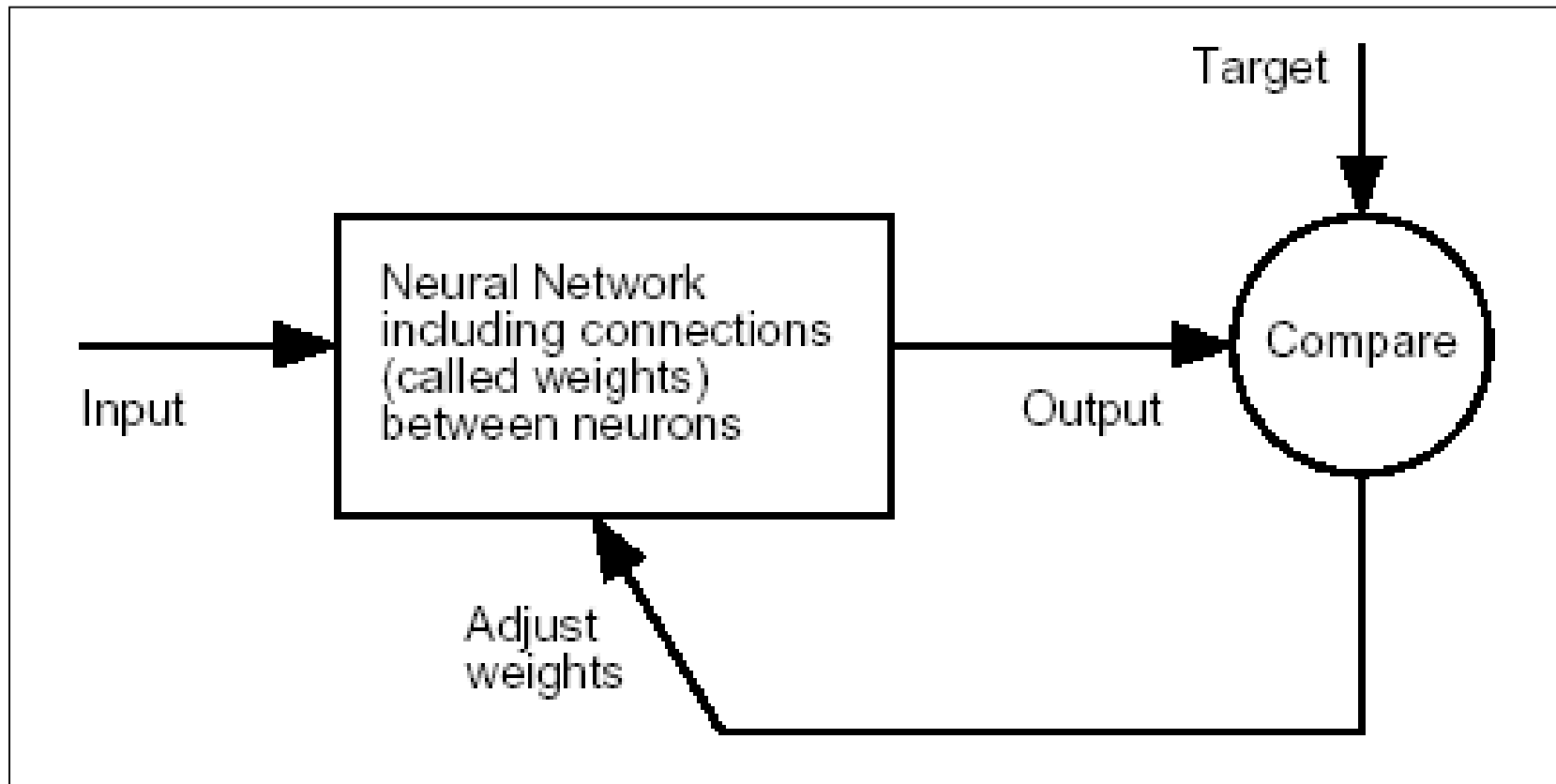
$$\Delta w_{ij} = \alpha \cdot (d_i - o_i) \cdot o_j$$

gdzie d_i jest pożądaną aktywacją, której wartość jest podawana przez nauczyciela.

Reguła ta jest nazywana regułą Widrowa-Hoffa (1960) lub regułą delty.

wagi zmieniane są proporcjonalnie do wielkości sygnałów wejściowych

Metody uczenia



Graficzna ilustracja procesu uczenia nadzorowanego

Metody uczenia

- Uczenie sieci, to proces minimalizacji błędu kwadratowego:

$$E[\vec{w}] = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

D = dane treningowe

Uczenie jest zbieżne, jeśli:

- ... nie ma nieliniowych zależności między danymi,
- ... stała uczenia α ma małą wartość,
- ... nie ma warstw ukrytych.

Budowa modelu

Co to znaczy zbudować model ?

Input: $X_1 X_2 X_3$ Output: Y Model: $Y = f(X_1 , X_2 , X_3)$

W przypadku SSN : matematyczny zapis $f(\dots)$ jest zbyt skomplikowany.

Jednakże wystarczająco charakteryzują go:

- liczba neuronów wejściowych,
- liczba warstw ukrytych,
- liczba neuronów w warstwach ukrytych,
- liczba neuronów wyjściowych,
- **wagi** wszystkich połączeń.

Budowa modelu SSN = określenie wartości powyższych parametrów

Ten i sześć następnych slajdów - autor: Angshuman Saha

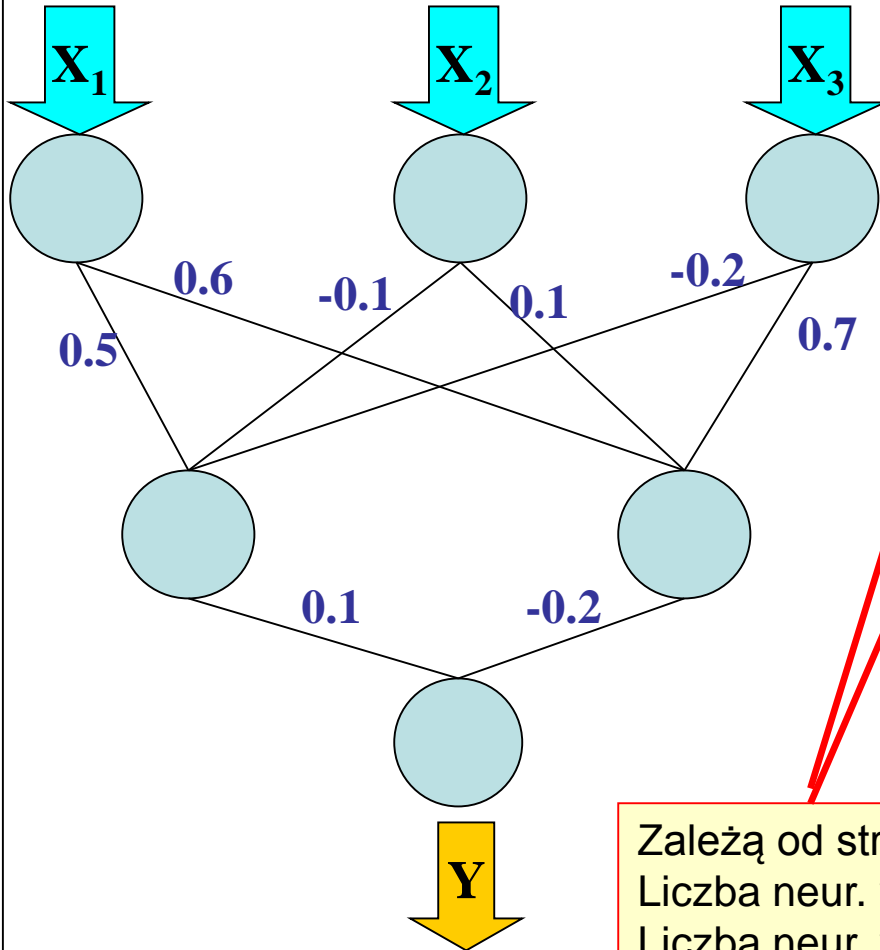
<http://www.geocities.com/adotsaha/NNinExcel.html>

Budowa modelu

Input: X_1 X_2 X_3

Output: Y

Model: $Y = f(X_1, X_2, X_3)$



Parametry	Przykład
L. neuronów wejściowych	3
L. warstw ukrytych	1
Neurony w warstwie ukr.	2
L. neuronów wyjściowych	1
Wagi	rysunek

Zależą od struktury problemu
 Liczba neur. wej. = liczba X
 Liczba neur. wyj. = liczba Y

Wolne parametry

Sztuczne sieci neuronowe

Budowa modelu

Input: $X_1 X_2 X_3$

Output: Y

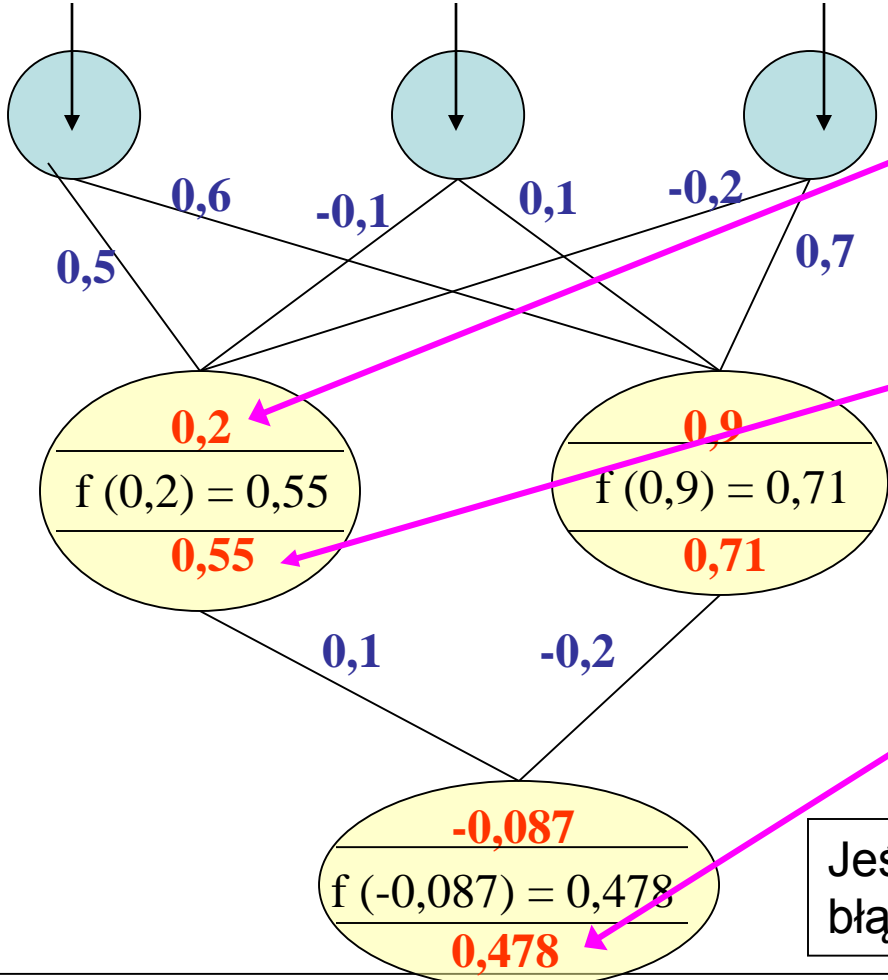
Model: $Y = f(X_1, X_2, X_3)$

$X_1 = 1$

$X_2 = -1$

$X_3 = 2$

$$0,2 = 0,5 * 1 - 0,1 * (-1) - 0,2 * 2$$



$$f(x) = 1 / (1 + e^{-x})$$

$$f(0,2) = 1 / (1 + e^{-0,2}) = 0,55$$

Wyliczone $Y = 0,478$

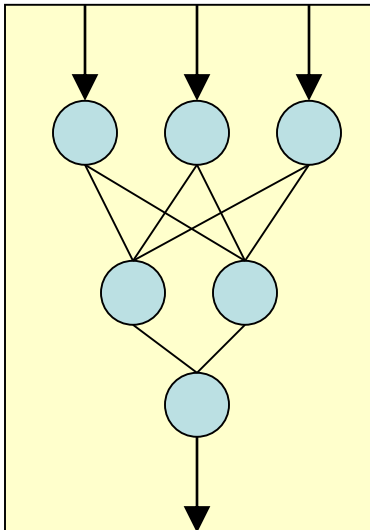
Jeśli wzorcowe $Y = 2$, to błąd oszacowania = $(2 - 0,478) = 1,522$

Budowa modelu

Input: $X_1 X_2 X_3$ Output: Y Model: $Y = f(X_1, X_2, X_3)$

L. neuronów wej. = liczba $X_i = 3$ L. neuronów wyj. = liczba $Y = 1$

Liczba warstw ukrytych = ??? próbuj **1** Nie ma reguły.
Liczba neuronów w warstwie = ??? próbuj **2** Metoda prób i błędów.



Mamy architekturę ... A co z wagami ???

Przy tej architekturze mamy do określenia 8 wartości wag:

$$\underline{W} = (w_1, w_2, \dots, w_8)$$

Dane treningowe: $(Y_i, X_{1i}, X_{2i}, X_{3i}) \quad i = 1, 2, \dots, n$

Dla określonych wag \underline{W} , otrzymamy oszacowane n wartości Y
 (V_1, V_2, \dots, V_n)

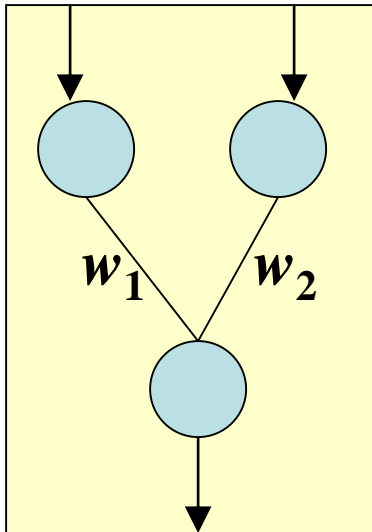
Oczywiście są one **funkcją** \underline{W} , czyli ogólna reguła brzmi:

Dobierz tak \underline{W} , by błąd oszacowania E był zminimalizowany

$$E = \sum_i (Y_i - V_i)^2$$

Uczenie sieci

Przykład prostej sieci z 2 neuronami wejściowymi i 1 - wyjściowym.



$$E(w_1, w_2) = \sum [Y_i - V_i(w_1, w_2)]^2$$

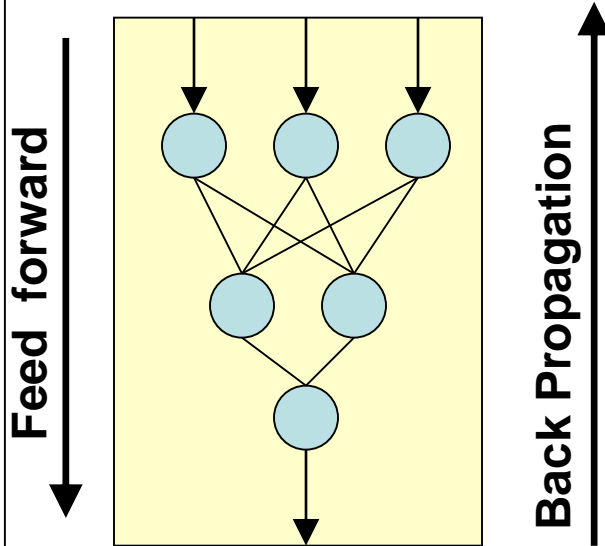
- Para (w_1, w_2) jest punktem w 2-wymiarowej płaszczyźnie.
- Dla każdej takiej pary mamy wartość E .
- Funkcja $E=f(w_1, w_2)$ wyznacza - ‘**przestrzeń błędów**’
- Cel: znaleźć taką parę wag, dla której E jest minimalny.

Algorytm hill climbing

1. Zaczynij od losowej pary (w_1, w_2)
2. Przesuń się do pary (w'_1, w'_2) , dla której błąd jest mniejszy.
3. Kontynuuj krok 2 aż dojdiesz do pary (w^*_1, w^*_2) , gdzie E jest minimalny.

Uczenie sieci

Algorytm wstecznej propagacji błędów Backpropagation Algorithm
uogólnienie reguły delty dla złożonych sieci



$$E = \sum (Y_i - V_i)^2$$

1. Zaczynij z wylosowanymi wartościami wag.
2. Zaprezentuj sieci pierwszą obserwację $(X_1, X_2, X_3) \rightarrow$ sieć $\rightarrow V_1$; $E = (Y_1 - V_1)^2$
3. Zmodyfikuj wagi tak, by zminimalizować błąd (sieć dopasowuje się do 1. obserwacji)
4. Zaprezentuj sieci kolejną obserwację. Zmodyfikuj wagi tak, by zminimalizować błąd.
5. Powtarzaj krok 3. aż do ostatniej obserwacji.
6. To kończy jeden cykl treningowy.
7. Wykonaj wiele takich cykli (epok), aż E osiągnie małą (minimalną?) wartość.

Uczenie sieci

Parametry uczenia:

- stała uczenia
- bezwładność (momentum)

Im wyższy współczynnik uczenia tym sieć uczy się szybciej, ale może wpaść w oscylacje

$$\Delta w_{ji}(t) = \alpha \delta_j(t) o_i(t) + \gamma \Delta w_{ji}(t-1)$$

$$\delta_j = -\frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} = -\frac{\partial E}{\partial o_j} f'(net_j)$$

Momentum - pozwala usunąć szybkie oscylacje

Sposób prezentacji danych:

- losowa prezentacja – element stochastyczny, uczenie on-line.
- ustalona kolejność.

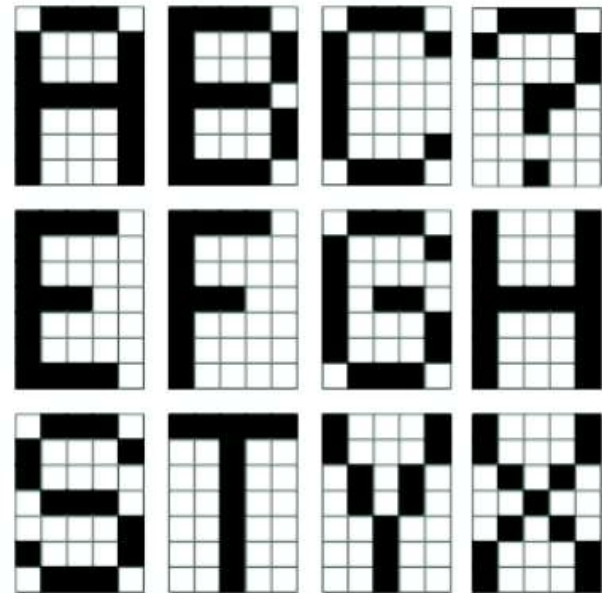
Sposób modyfikacji wag (wsadowy vs. przyrostowy).

Metody uczenia

- **Sposób modyfikacji wag**
- **Uczenie przyrostowe**, w tej odmianie minimalizuje się funkcję błędów dla każdej pary wektorów zmiennych wejściowych i wzorców osobno; modyfikacja wag następuje każdorazowo po takiej operacji.
- **Uczenie wsadowe**, w tej odmianie minimalizuje się błąd średniokwadratowy wyznaczony dla wszystkich próbek ze zbioru uczącego, również korekta wag następuje po zaprezentowaniu całego zbioru uczącego.

Uczenie sieci – przykład*

- Uczenie metodą propagacji wstecznej jednowarstwowej sieci klasyfikującej dwanaście znaków, reprezentujących odrębne klasy (zajmują wierzchołki 35-wymiarowej kostki i są liniowo separowalne, a tym samym nauczona sieć jednowarstwowa powinna je poprawnie klasyfikować).
- Wyniki klasyfikacji odczytywano przyporządkowując sygnałom na wyjściach neuronów liczbę 1, gdy przekraczają wartość 0,5, a liczbę 0 w przeciwnym razie. Badana sieć zawierała 12 neuronów o logistycznych funkcjach aktywacji z parametrami $\beta = 1$.

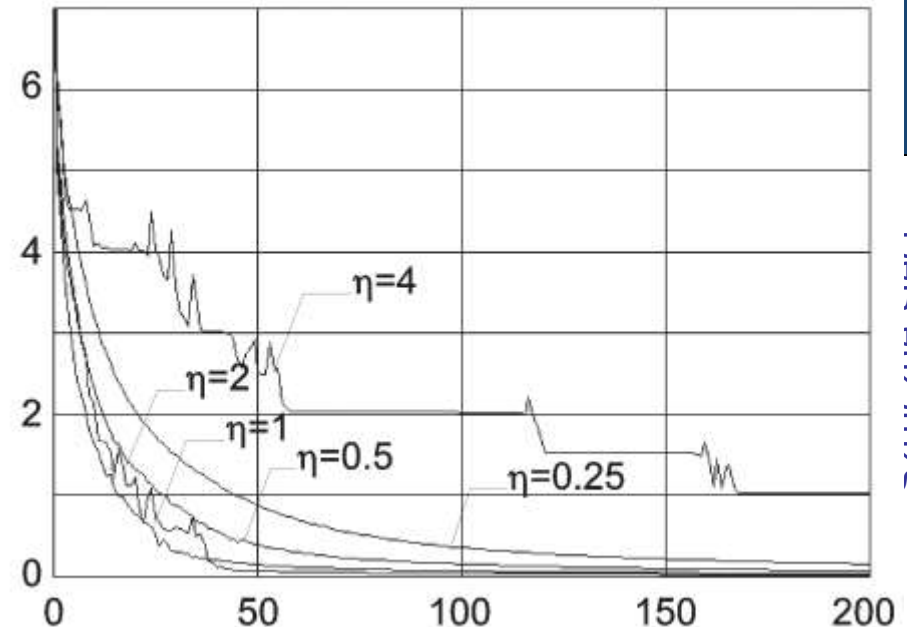


*Za: W. Jędruch, Sztuczna inteligencja, Politechnika Gdańska, 2010.

Uczenie sieci – przykład

- Przeprowadzono 200 cykli uczenia sieci dla kilku wartości współczynnika η , po których uzyskano następujące błędy łączne:
 $\eta = 4, E = 1,0040$ $\eta = 2, E = 0,0138$
 $\eta = 1, E = 0,0294$ $\eta = 0,5, E = 0,0651$ $\eta = 0,25, E = 0,1469$

Z rysunku widać wpływ współczynnika η na szybkość uczenia. Zbyt duża wartość ($\eta=4$) powoduje duże oscylacje i przeregulowania i w efekcie duży błąd końcowy. Pośrednie wartości ($\eta=2$ lub 1) powodują, pomimo niewielkich oscylacji, szybkie osiągnięcie małych wartości błędu. Wreszcie zbyt małe wartości ($\eta=0,5$ lub $0,25$) powodują systematyczne i bez oscylacji, ale powolne malenie błędu.



Uczenie sieci – przykład

- Dla ilustracji reakcji sieci poniżej przedstawiono sygnały wyjściowe neuronów nauczonej sieci (dla $\eta = 2$) odpowiadającej na przykładowy obraz wejściowy (znak A), gdzie w nawiasach przy liczbach podano znaki reprezentowane przez poszczególne neurony:

(A)	0,9799,	(B)	0,0039,	(C)	0,0113,	(?)	0,0103
(E)	0,0001,	(F)	0,0100,	(G)	0,0020,	(H)	0,0123
(S)	0,0002,	(T)	0,0014,	(Y)	0,0001,	(F)	0,0086

Zalety SSN

1. Rozwiązuje problemy, które klasyczne algorytmy nie są w stanie rozwiązać, mówimy tutaj przede wszystkim o zadaniach, w których skład wchodzi kojarzenie czy przewidywanie.
2. Nie potrzebują specjalnego algorytmu, całym zadaniem jest zaprojektowanie odpowiedniej struktury sieci odpowiadającej problemowi oraz pokierowanie procesem uczenia się sieci. Niepotrzebne są charakterystyki modelu, założenia, ograniczenia. Wystarczy podać cel i dużo przykładów praktycznego osiągnięcia tego celu. Wprawdzie sieć nie poda reguł kierujących określonym zjawiskiem, ale poda rozwiązanie problemu.

Zalety SSN

3. Działają jako całość, każdy element ma pewien wkład w realizację zadania, dlatego też mogą działać pomimo błędów (np. struktury czy też błędów w danych). Klasyczny algorytm w sytuacji błędu np. w strukturze algorytmu lub w zestawie danych najczęściej „wywraca się”, uniemożliwiając działanie programu. Natomiast SSN dopiero w przypadku ilości błędów powyżej pewnego granicznego poziomu nie jest w stanie poprawnie działać.
4. Umiejętność uogólniania wiedzy. *Znaczy to dokładnie tyle, że jeśli sieć nauczy się, powiedzmy rozpoznawać kolory: czerwony i żółty, to rozpozna również różowy i bladożółty, czyli kolory podobne do znanych.*

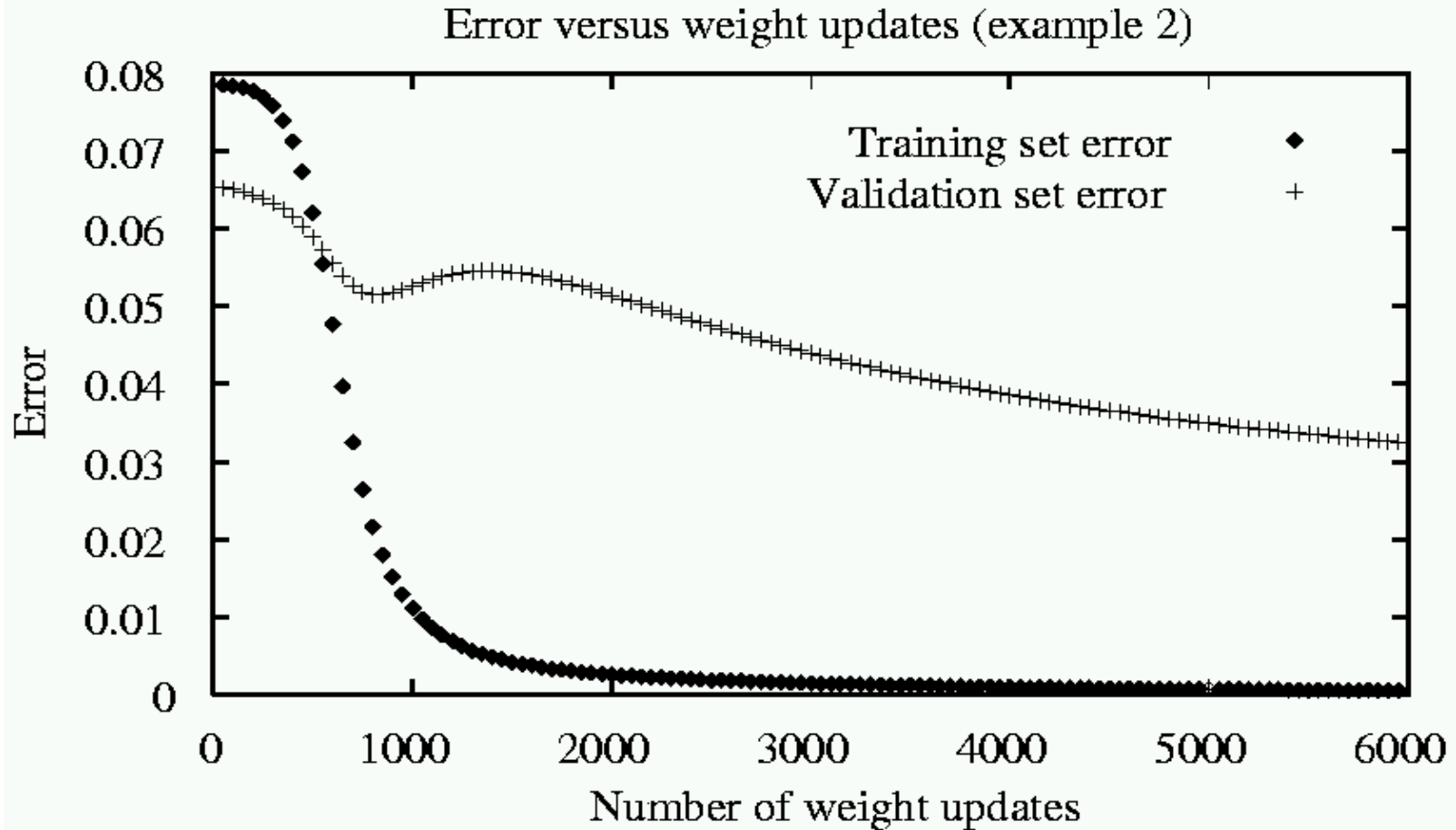
Korbicz J., Obuchowicz A., Uciński D. - "Sztuczne sieci neuronowe. Podstawy i zastosowania", Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994

Wady SSN

1. Operują pojęciami rozmytymi, nie określają dokładnie, lecz w przybliżeniu, podobnie zresztą jak ludzki mózg. Dlatego nie mogą być stosowane tam, gdzie potrzebna jest precyzyjna wartość.
2. Nie radzą sobie z rozumowaniem wieloetapowym, gdy z jednych wniosków trzeba wysnuć inne, a z tych - kolejne itd. Wynika to z faktu, że sieć neuronowa działa jednoetapowo, tylko w jednym kroku.
3. Proces uczenia się sieci może trwać bardzo długo; wszystko zależy od stopnia skomplikowania sieci oraz złożoności problemu do rozwiązania.
4. Trudno wysuwać wnioski co do istoty zależności między danymi wyjściowymi a wejściowymi: sztuczna sieć neuronowa działa jak „czarna skrzynka”, do której wrzucamy dane i otrzymujemy wyniki, natomiast logiki systemu i schematów myślenia nie jesteśmy w stanie odkryć.
5. Łatwo „przeuczyć sieć”, tzn. sieć nauczy się na pamięć podawanych przykładów w fazie uczenia i w momencie testowania na innych przykładach jest bezużyteczna, nie potrafi właściwie rozwiązać zadania (nie ma zdolności do generalizacji).

Wady SSN

- **Uczenie sieci - przetrenowanie**



Cykl życia SSN

1. Określenie zmiennej wyjściowej oraz zmiennych wejściowych.
2. Gromadzenie danych.
3. Wstępne przetwarzanie danych (preprocessing).
4. Podział zbioru danych.
5. Dobór odpowiedniej architektury sieci.
6. Uczenie sieci.
7. Testowanie i weryfikacja.
8. Zastosowanie sieci.

Topologia SSN

- Ze względu na rodzaj powiązań między elementami sieci wyróżniamy:
 - *sieci jednokierunkowe* (feed-forward), w których dane przepływają w jednym kierunku od elementów wejściowych do wyjściowych.
 - *sieci rekurencyjne* zawierające połączenia ze sprzężeniami zwrotnymi.

Zastosowania SSN

- Większość zastosowań SSN w zarządzaniu i ekonomii polega na wykorzystaniu sieci jednokierunkowych, wielowarstwowych do klasyfikacji, rozpoznawania wzorców i predykcji oraz sieci Hopfielda i Kohonena do celów optymalizacyjnych.
- **Marketing**
 - SSN jest elementem systemu ekspertowego do śledzenia i rezerwowania miejsc w samolotach.
 - SSN może służyć do klasyfikowania klientów np. pod względem ryzyka handlowego.
 - Prognozowanie sprzedaży.

Zastosowania SSN

- Harmonogramowanie i planowanie
 - Optymalizacja struktury systemu przesyłania energii.
 - Optymalizacja rozkroju materiału.
 - Optymalizacja struktury sieci satelitów komunikacyjnych.
 - Sterowanie ruchem samolotów.
- Kontrola jakości
 - Sieć może szacować wpływ różnych zmiennych i ich kombinacji na jakość produktu (ćwiczenie).
 - Montaż głowic dysków jest diagnozowany przez sieć firmy Nestor. Zwiększenie lub zmniejszenie wypukłości wycinanych części jest wykrywane, a moment, w którym tolerancja przestaje być dotrzymywana, jest sygnalizowany.

Zastosowania SSN

- Diagnostyka
 - koncern Ford Motor Company wdrożył system diagnostyczny silników.
- Bezpieczeństwo
 - Firma SIAC oferuje system monitorowania bagaży na lotniskach (stosowany jest np. na lotnisku JFK w Nowym Jorku).
 - Firma Nestor sprzedaje system do rozpoznawania podpisów mający duże zastosowanie w bankach; sieć działa w czasie rzeczywistym z efektywnością 92-98%.

Zastosowania SSN

- Prognozowanie
 - Prognoza rozwoju rynku energetycznego [Werbos].
 - Prognoza zapotrzebowania na energię elektryczną.
- Rynek akcji
 - Prognoza akcji IBM (nieudana, White 1988).
 - Kursy wymiany walut.