

Inteligentne systemy decyzyjne: Uczenie maszynowe – sztuczne sieci neuronowe

wykład 1.

- Właściwości sieci neuronowych
- Model matematyczny sztucznego neuronu
- Rodzaje sieci neuronowych
- Przegląd głównych zastosowań

Dr inż. Paweł Żwan
Katedra Systemów Multimedialnych
Politechnika Gdańska

Cechy sieci neuronowych

Modelowanie liniowe a modelowanie nieliniowe:

- Modele liniowe (aproksymacja liniowa) często nie sprawdzają się prowadząc do zbyt szybko wyciąganych wniosków o “niemożności” matematycznego opisu danego systemu.
- **Zdolność sieci do odwzorowywania** nadzwyczaj złożonych **funkcji nieliniowych**, co umożliwia ich szerokie zastosowania
- Tworzenie modeli przy pomocy sieci neuronowych jest najwygodniejszym rozwiązaniem problemu.
- **Niezależność** złożoności algorytmu **od wielowymiarowości** (łatwość aproksymacji funkcji nieliniowych z dużą ilością zmiennych)

Cechy sieci neuronowych

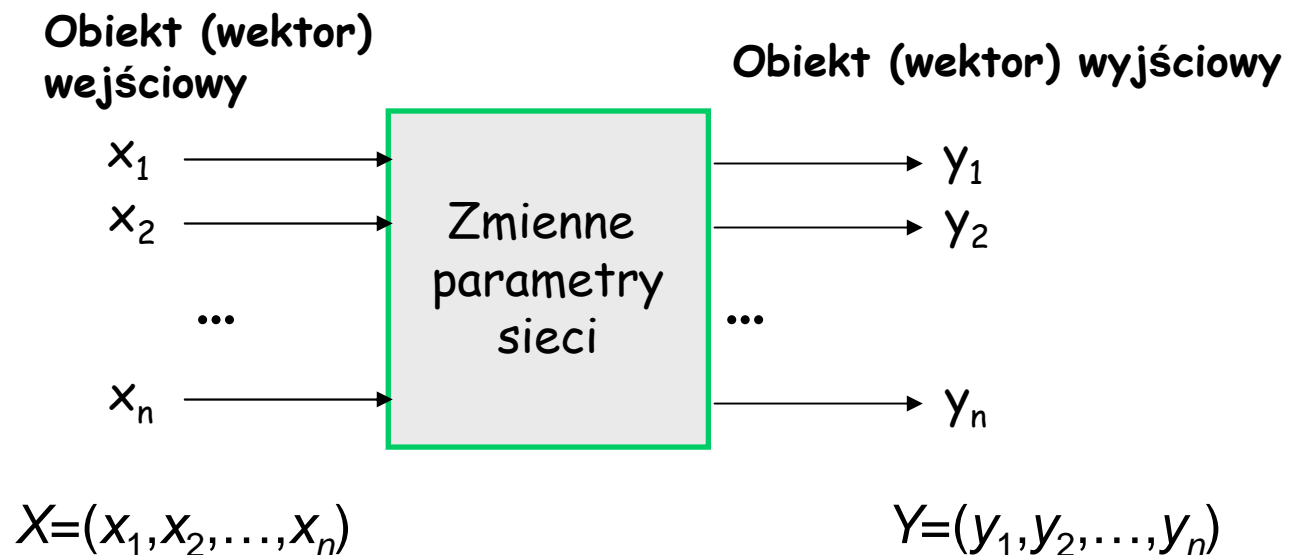
- Samodzielność konstruowania modeli potrzebnych użytkownikowi bez znajomości „a priori” przepisu na działanie - automatyczne **uczenie się** na podanych przez niego przykładach.
- W celu zaprojektowania sieci neuronowej należy:
 - zgromadzić reprezentatywne dla problemu dane
 - przykłady praktyczne
 - Uruchomić algorytmu uczenia w celu wytworzenia w pamięci sieci potrzebnego modelu.
- Działanie sieci musi realizować wszystkie potrzebne funkcje związane z działaniem wytworzonego modelu dla danych, które nie musiały być obecne podczas treningu systemu

Przygotowanie danych

- Sieci neuronowe wymagają od użytkownika:
 - wiedzy empirycznej dotyczącej wyboru i przygotowania danych uczących:
 - a) dane uczące (ang. *training data*) – do treningu
 - b) dane weryfikujące (ang. *validation data*) – do treningu
 - c) dane testowe (ang. *test data*) – do testowania po treningu
 - wyboru właściwej architektury (struktury sieci)
 - lecz... **nie wymagają posiadania szczególnie specjalistycznej wiedzy teoretycznej niezbędnej do zbudowania modelu matematycznego.**
- Sieć buduje model sama na podstawie nauki.
- Poziom wiedzy teoretycznej niezbędnej do skutecznego zbudowania modelu jest przy stosowaniu sieci neuronowych znacznie niższy niż w przypadku stosowania tradycyjnych metod statystycznych.

Automatyczne tworzenie modelu matematycznego

- Sieci neuronowe automatycznie tworzą model matematyczny dla danego zagadnienia na podstawie danych uczących
- W wyniku procesu uczenia umieją odwzorować różne złożone zależności pomiędzy sygnałami wejściowymi i wyjściowymi.

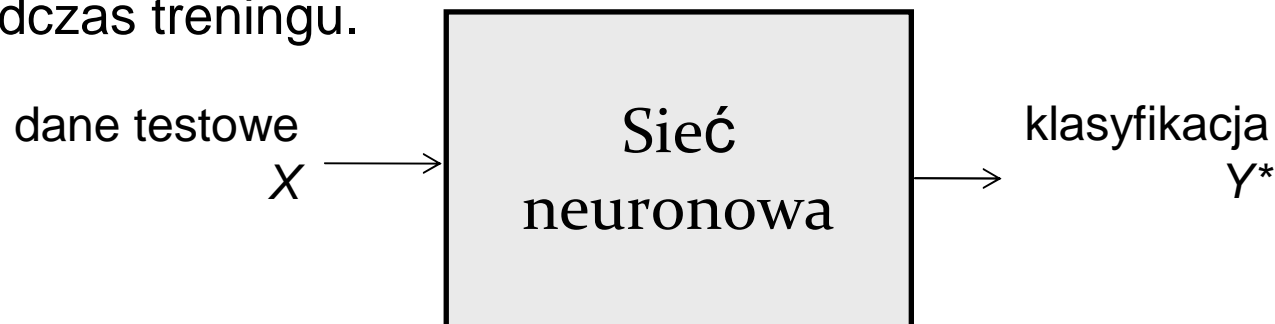


Sieć neuronowa jako „czarna skrzynka” rozwiązująca dany problem

- Do nauki sieci neuronowej potrzebne są wektory (obiekty) wejściowe x_i wraz z właściwymi im obiektami (wektorami) wyjściowymi y_i



- Aplikacja (działanie sieci) związana jest z generowaniem przez sieć wektorów wyjściowych y_i na podstawie wektorów x_i , które nie były użyte podczas treningu.



Sieć neuronowa a ludzki mózg

- Początek dziedziny – praca:

W. S. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, No 5, 1943, pp. 115-133.

która zawiera pierwszy matematyczny opis komórki nerwowej i powiązanie tego opisu z problemem przetwarzania danych.

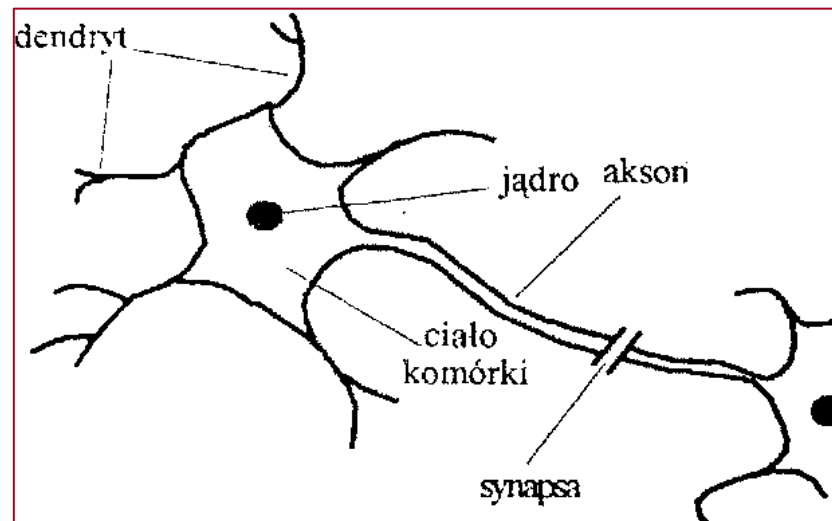
- Oparcie działania sieci na prostym modelu, przedstawiającym wyłącznie najbardziej podstawową istotę działania biologicznego systemu nerwowego.
- Naśladownictwo (w pewnym, bardzo ograniczonym zakresie działania ludzkiego mózgu).
- Odporność na uszkodzenia nawet znacznej części elementów (właściwość biologicznych systemów nerwowych)

Sieć neuronowa a ludzki mózg

Mózg to bardzo duża (ok. 10 miliardów) liczba elementarnych komórek nerwowych czyli neuronów połączonych w formie skomplikowanej sieci.

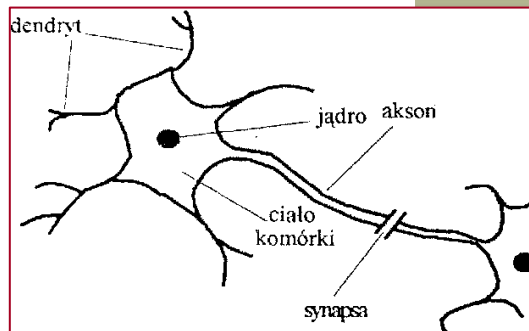
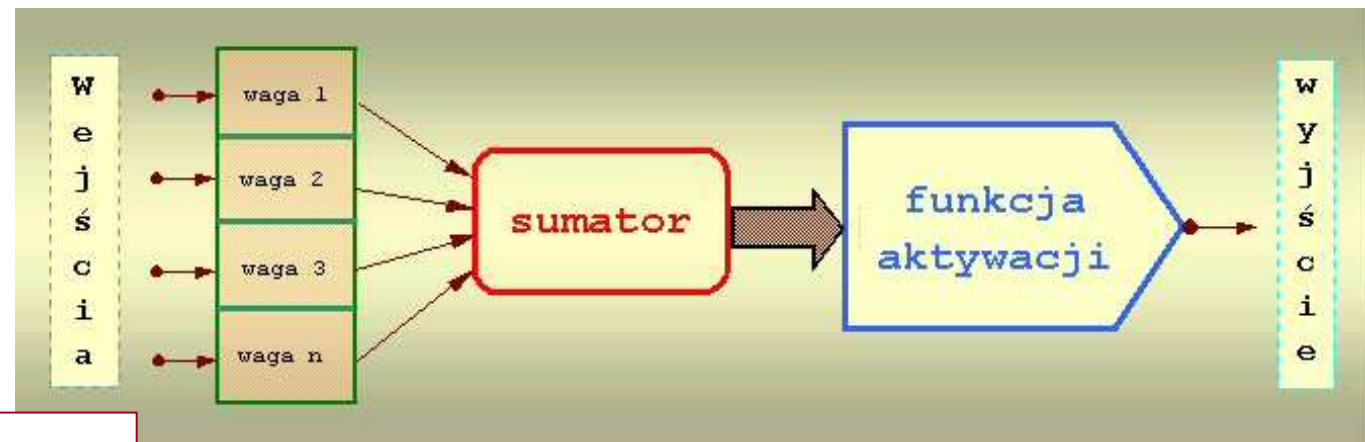
Średnio na jeden neuron przypada kilka tysięcy połączeń, ale dla poszczególnych komórek ilości połączeń mogą się między sobą różnić.

Biologiczny neuron:

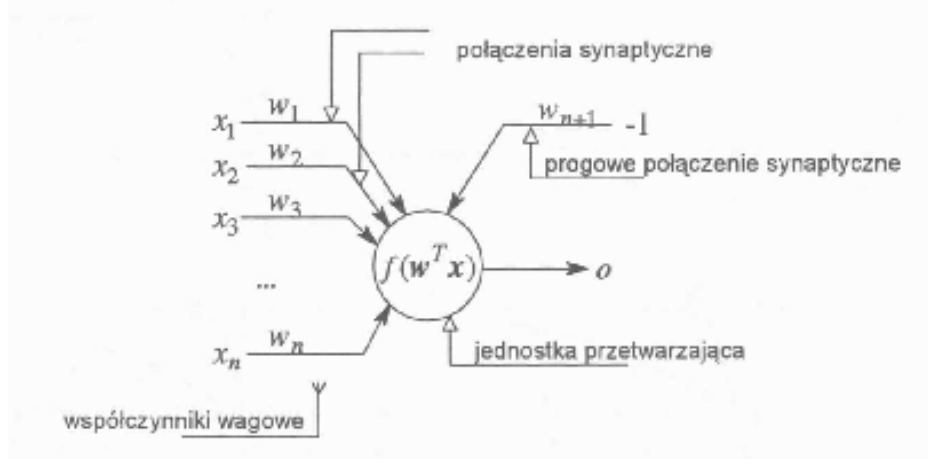


Sieć neuronowa a ludzki mózg

* Biologiczny neuron zostaje zamieniony na model matematyczny, i z takich „cegiełek” budowane są skomplikowane struktury decyzyjne naśladujące pewne funkcjonalności ludzkiego mózgu.



Model matematyczny sztucznego neuronu



Wartość wyjściowa neuronu o jest określana w oparciu o wzór: $o=f(\mathbf{w}^T\mathbf{x})$,

gdzie: \mathbf{w} – wektor wag połączeń wejściowych

\mathbf{x} – wektor wartości sygnałów wejściowych

f – funkcja aktywacji

Dodatkową wagą jest waga progowa, dlatego wektory \mathbf{w} i \mathbf{x} określone są jako:

$$\mathbf{w}=[w_1, w_2, \dots, w_n, w_{n+1}], \mathbf{x}=[x_1, x_2, \dots, x_n, -1]$$

Funkcje aktywacji neuronów

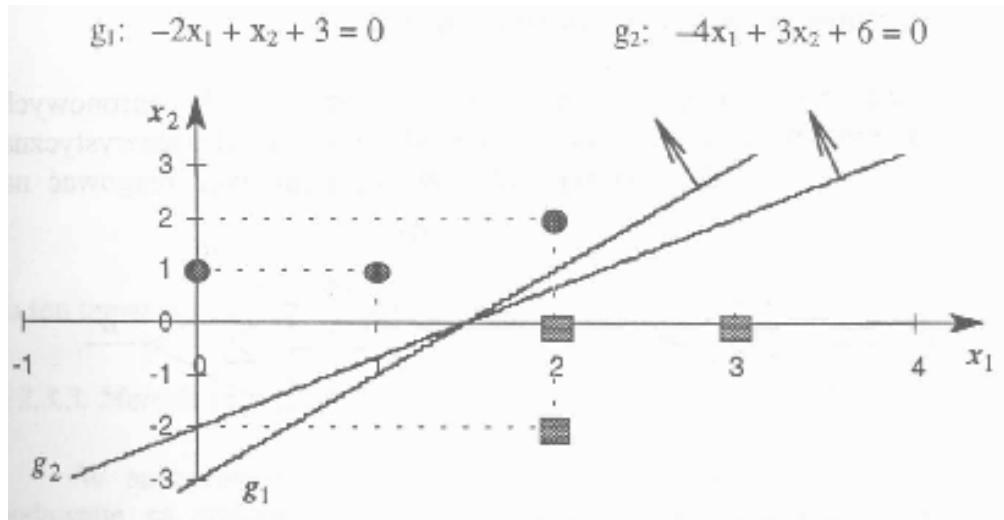
funkcja	Wzór funkcji	Wzór pochodnej
Sigmoida	$f(x) = \frac{1}{1 + e^{-\beta x}}$	$f'(x) = \beta * (1 - f(x)) * f(x)$
Tangens hiperboliczny	$f(x) = \tanh(\beta * x)$	$f'(x) = \beta(1 - f^2(x))$
Sinusoida	$f(x) = \sin(\beta * x)$	$f'(x) = \beta \sqrt{1 - f^2(x)}$
Cosinusoida	$f(x) = \cos(\beta * x)$	$f'(x) = -\beta \sqrt{1 - f^2(x)}$
$\frac{x}{(1 + x)}$ (bez nazwy)	$f(x) = \frac{\beta * x}{(1 + \beta * x)}$	$f'(x) = \frac{\beta}{1 + \beta * x } - \frac{ \beta * x }{(1 + \beta * x)^2}$

Wymagane cechy funkcji aktywacji to:

- ciągłe przejście pomiędzy swoją wartością maksymalną a minimalną (np. 0-1),
- łatwa do obliczenia i ciągła pochodna,
- możliwość wprowadzenia do argumentu parametru beta do ustalania kształtu krzywej.

Naprostrzy klasyfikator neuronowy – dyskretny dychotomizator

- klasyfikacja n – wymiarowego obiektu do jednej z dwóch klas
- może składać się tylko z jednego neuronu
- Gdy wyjście neuronu $o \geq 0$ – klasyfikacja do klasy 1. , gdy wyjście neuronu $o < 0$ – klasyfikacja do klasy 2.



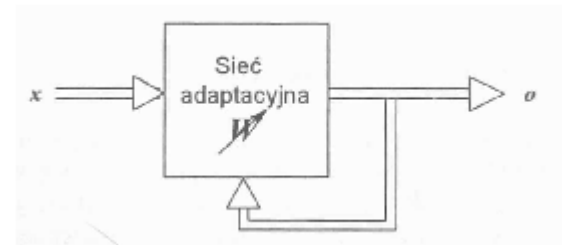
$$g_1: -2x_1 + x_2 + 3 = 0$$

$$g_2: -4x_1 + 3x_2 + 6 = 0$$

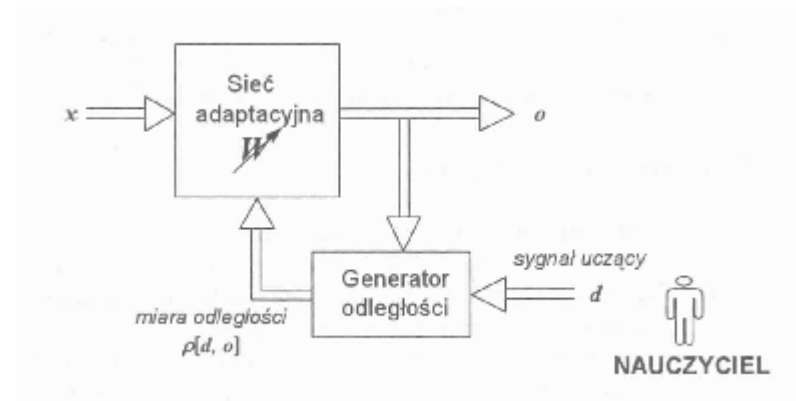
Podział sieci neuronowych

- Podział ze względu na metodę treningu:

- **trening bez nadzoru**

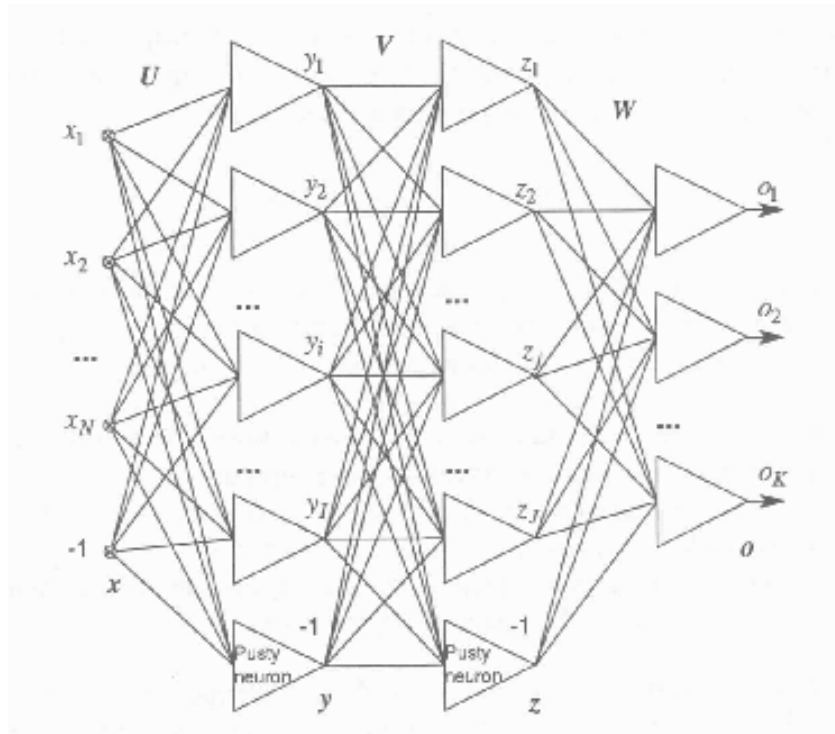


- **trening z nadzorem**



Podział sieci neuronowych

- Podział ze względu na strukturę:
 - **sieci jednokierunkowe**



$\mathbf{x}=[x_1, x_2, \dots, x_N, -1]$ – wektor wejściowy

$\mathbf{y}=[y_1, y_2, \dots, y_I, -1]$ – wektor wyjściowy pierwszej warstwy ukrytej

$\mathbf{z}=[z_1, z_2, \dots, z_J, -1]$ – wektor wyjściowy drugiej warstwy ukrytej

$\mathbf{o}=[o_1, o_2, \dots, o_K, -1]$ – wektor wyjściowy

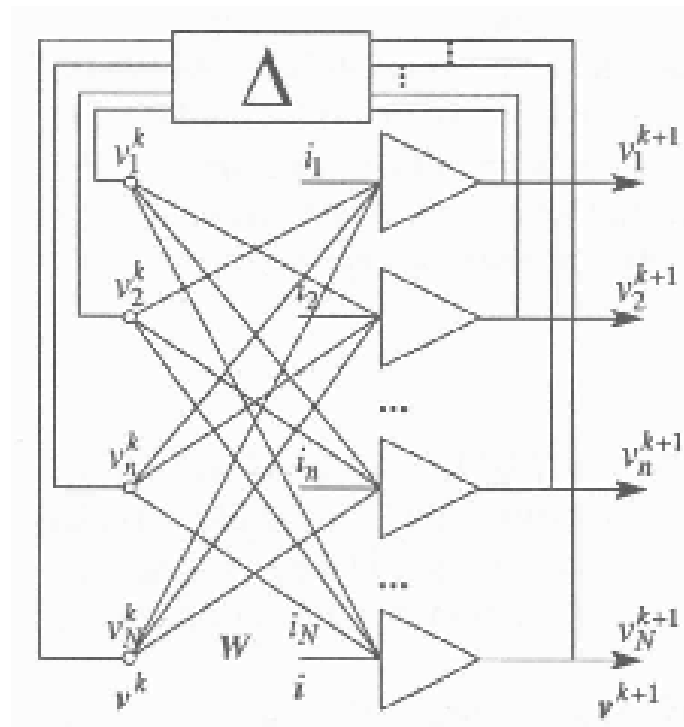
Macierze $\mathbf{U}, \mathbf{V}, \mathbf{W}$ zawierają Współczynniki wagowe dla wszystkich Połączeń synaptycznych

Podział sieci neuronowych

- **sieci ze sprzężeniem zwrotnym**

- * wejście przynajmniej jednego neuronu jest połączone bezpośrednio lub pośrednio z jego wejściem

- * przykład:
sieć Hopfielda

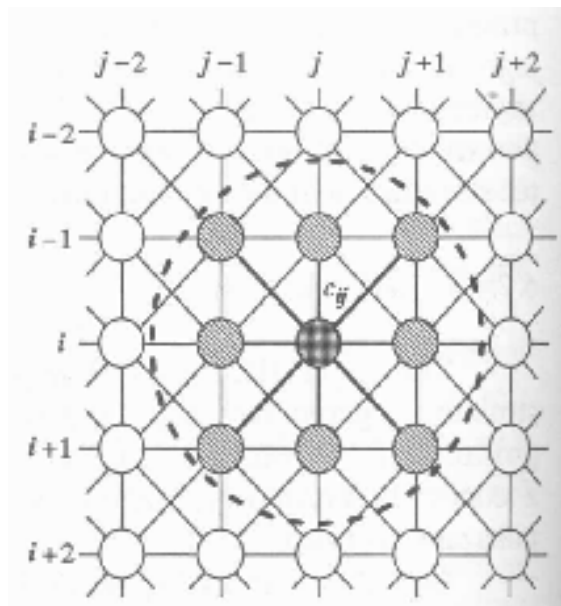


Podział sieci neuronowych

- Sieci komórkowe

* topologia oparta na dowolnej regularnej strukturze geometrycznej

* przykład: topologia płaskiej siatki prostokątnej:



- Neurony zgrupowane w I wierszach i J kolumnach
- Dowolna komórka połączona jest tylko z neuronami w najbliższym sąsiedztwie (konieczna def. promienia sąsiedztwa – np.1)
- Zastosowanie: Przetwarzanie obrazów dwuwymiarowych

Przykłady zastosowań sieci neuronowych

Funkcje pełnione przez sieć można ująć w kilka podstawowych grup:

- **aproksymacji i interpolacji**
- **rozpoznawania i klasyfikacji wzorców**
- **kompresji**
- **predykcji i sterowania**
- **asocjacji**

Sieć neuronowa pełni w każdym z tych zastosowań rolę **uniwersalnego aproksymatora** funkcji wielu zmiennych, realizując **funkcję nieliniową** o postaci $\mathbf{y} = \mathbf{f}(\mathbf{x})$, gdzie \mathbf{x} jest wektorem wejściowym, a \mathbf{y} realizowaną funkcją wektorową wielu zmiennych.

Duża liczba zadań modelowania, identyfikacji, przetwarzania sygnałów da się sprowadzić do zagadnienia aproksymacyjnego.

Przegląd zastosowań sieci

Przy **klasyfikacji i rozpoznawaniu wzorców** sieć uczy się podstawowych cech tych wzorców, takich jak odwzorowanie geometryczne układu pikselowego wzorca, rozkładu składników głównych wzorca, czy jego innych parametrów. Dobre uczenie polega na podawaniu wzorców o dużych różnicach, stanowiących podstawę podjęcia decyzji przypisania ich do odpowiedniej klasy.

Przy **predykcji** zadaniem sieci jest określenie przyszłych odpowiedzi systemu na podstawie ciągu wartości z przeszłości. Mając informacje o wartościach zmiennej x w chwilach poprzedzających predykcję $x(k-1)$, $x(k-2)$,, $x(k-N)$, sieć podejmuje decyzje, jaka będzie estymowana wartość $x(k)$ badanego ciągu w chwili aktualnej k .

Przegląd zastosowań sieci

W zagadnieniach **identyfikacji i sterowania procesami dynamicznymi** sieć neuronowa pełni zwykle kilka funkcji. Stanowi model nieliniowy tego procesu, pozwalający na wypracowanie odpowiedniego sygnału sterującego. Pełni również funkcje układu śledzącego i nadążnego, adaptując się do warunków środowiskowych – w tej dziedzinie najczęściej stosuje się **sieci ze sprzężeniem zwrotnym**.

W zadaniach **asocjacji** sieć neuronowa pełni rolę pamięci skojarzeniowej. Można wyróżnić pamięć asocjacyjną, w przypadku której skojarzenie dotyczy tylko poszczególnych składowych wektora wejściowego oraz pamięć heteroasocjacyjną, gdzie zadaniem sieci jest skojarzenie ze sobą dwóch wektorów. Jeśli na wejście sieci podany będzie wektor odkształcony (np. o elementach zniekształconych szumem bądź pozbawiony pewnych elementów danych), sieć neuronowa jest w stanie odtworzyć wektor oryginalny, pozbawiony szumów, generując przy tym pełną postać wektora stowarzyszonego z nim.

cdn... czyli ... plan wykładu drugiego

- Uczenie sieci jednokierunkowych – reguła delta i algorytm wstecznej propagacji błędu
- Metody poprawy szybkości nauki sieci – dobór parametrów nauki
- Dobór optymalnej architektury sieci
- Generalizacja, a zapamiętywanie sieci – metody kontroli zdolności generalizacyjnych sieci