

www.kwmimkm.polisi.pl

METODY HEURYSTYCZNE

wykład 4

www.kwmimkm.polisi.pl

OCENA DZIAŁANIA AE

2

www.kwmimkm.polisi.pl

LOSOWOŚĆ W AE

Różne zachowanie algorytmu w niezależnych uruchomieniach przy jednakowych ustawieniach parametrów i identycznych populacjach początkowych.

Losowość jest wprowadzana w AE:

- Podczas generowania populacji początkowej
- W procesie wyboru populacji potomnej na drodze reprodukcji (ukierunkowanie działania AE).
- Podczas działania operatorów ewolucyjnych (próbki w przestrzeni roboczej).

3

www.kwmimkm.polisi.pl

- Należy dokonać wielu niezależnych uruchomień dla losowej próby różnych populacji bazowych $P(0)$.
- W przypadku wielu uruchomień dla tej samej populacji $P(0)$ można mówić o właściwościach danego algorytmu dla konkretnej populacji początkowej.
- Porównywane dwa algorytmy: wskazane jest aby próba ta była taka sama (dla każdej losowej populacji początkowej uruchamia się dwa porównywane algorytmy).

4

www.kwmimkm.polisi.pl

ANALIZA STATYSTYCZNA:

- Analiza wartości oczekiwanej i odchylenia standardowego wartości przystosowania w populacji bazowej.
- Uwzględnienie informacji o min. i max. osiągniętej wartości.
- Uwzględnienie informacji o liczbie przypadków niewiele różnych od najlepszego.
- Prezentacja wyników w postaci histogramu pozwala ocenić właściwości rozkładu, które mogą umknąć przy analizie ograniczonej jedynie do statystyk liczbowych.

Przykładowy histogram:

5

www.kwmimkm.polisi.pl

KRZYWE ZBIEŻNOŚCI

- Wykresy zmian wartości rozwiązania roboczego (średniego, najlepszego, najgorszego) w czasie.
- Kreśli się je:
 - dla pojedynczego uruchomienia algorytmu;
 - dla wielu niezależnych uruchomień.
 (Uśrednione – bardziej reprezentatywne, lecz gubi się niektóre informacje o zachowaniu AE w pojedynczych uruchomieniach).

Przykładowe krzywe zbieżności (dla 1 uruchomienia algorytmu):

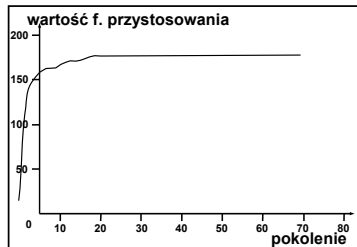
6

Szczególna krzywa zbieżności:

www.kwmimkm.polisi.pl

Wykres zmian w kolejnych pokoleniach wartości przystosowania najlepszego osobnika znalezionej od początku działania AE.

Po zakończeniu działania AE osobnik ten jest rozwiązaniem wyz. przez pojedyncze uruchomienie AE.



7

KRYTERIA ZATRZYMANIA AE

www.kwmimkm.polisi.pl

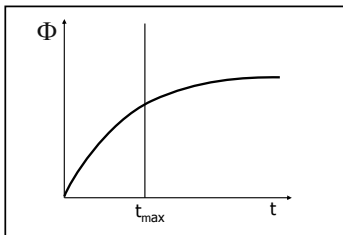


8

KRYTERIUM MAKSYMALNEGO KOSZTU

www.kwmimkm.polisi.pl

- Algorytm kończy działanie, jeśli koszt algorytmu przekroczy założoną wartość maksymalną K_{max} .
- Często przyjęta odmiana – przyjęcie pewnej maksymalnej dopuszczalnej liczby pokoleń algorytmu.



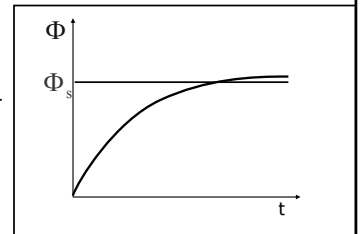
9

KRYTERIUM ZADOWALAJĄCEGO POZIOMU FUNKCJI PRZYSTOSOWANIA

www.kwmimkm.polisi.pl

Zatrzymanie działania gdy AE znajdzie rozwiązanie o wartości funkcji przystosowania określonej przez użytkownika jako zadowalająca Φ_s .

- ⊗ Zwykle nie jest łatwo (bez dostatecznie dobrej znajomości funkcji przystosowania) określić wartość zadowalającą.
- ⊗ AE może działać dowolnie długo (należy dodatkowo określić maksymalny koszt znalezienia rozwiązania).



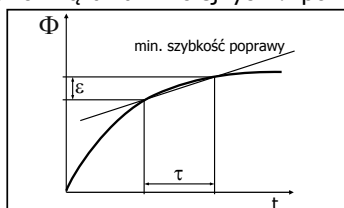
10

KRYTERIUM MIN. SZYBKOŚCI POPRAWY

www.kwmimkm.polisi.pl

Algorytm jest zatrzymywany, jeśli w kolejnych τ obliczeniach wartości funkcji przystosowania nie uda się poprawić wyniku o więcej niż ϵ .

Często $\epsilon = 0$ - algorytm zatrzymywany, jeśli nie uda się uzyskać lepszego rozwiązania w kolejnych τ pokoleniach.



11

SZTUCZNE SIECI NEURONOWE

www.kwmimkm.polisi.pl



12

HISTORIA SSN

www.kwmimkm.polisi.pl

- **Walter Pitts, Warren McCulloch** (1943) – opracowanie matematyczne pojęcia sztucznego neuronu. Udowodnili też, iż ich wynalazek jest w stanie odwzorować w swym działaniu dowolną funkcję logiczną.
- **Donald Olding Hebb** (1949) - zasada uczenia się Hebba (Hebbian learning) dla sztucznych sieci neuronowych (SSN).
- **Frank Rosenblatt** (1958) – pierwszy funkcjonujący model SSN (perceptron) oraz pierwszy z algorytmów uczenia SSN.



13

- **Bernard Widrow i Ted Hoff** (1960) - neuron typu Adaline z liniową funkcją aktywacji oraz algorytm uczenia LMS (Least Mean Square), zwany regułą delty lub regułą Widrowa-Hoffa.

- **John Hopfield** (1982) – teoria pamięci asocjacyjnej, jako istoty działania sieci rekurencyjnych (sieci Hopfielda).

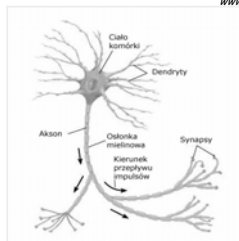
- **Paul Werbos** (1974); **David Rumelhart, Geoffrey Hinton, Ronald Williams** (1986) - wsteczna propagacja błędów (backpropagation) – pozwala na rozwiązywanie problemów liniowo nieseparowalnych.



14

Komórki nerwowe (neurony)

www.kwmimkm.polisi.pl



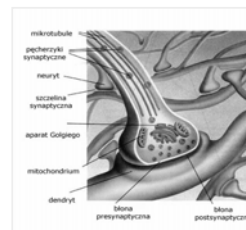
- **Dendryty** – zbierają sygnały z innych komórek nerwowych.
- **Ciało komórki** – agreguje sygnały wejściowe i tworzy sygnał wyjściowy.
- **Akson** – wyprowadza sygnał wyjściowy i przekazuje go dalej.



15

Synapsa - przekazuje sygnał między aksonem a dendrytem (każda komórka nerwowa posiada średnio kilka tysięcy synaps).

www.kwmimkm.polisi.pl



Chemiczno-elektryczne przekazywanie sygnałów:

- Pod wpływem przychodzących bodźców wydzielane są **neuroprzekaźniki**.
- Neuroprzekaźniki oddziałują na błonę komórki zmieniając jej potencjał elektryczny.



16

- Poszczególne synapsy różnią się wielkością oraz możliwością gromadzenia neuroprzekaźników w pobliżu błony synaptycznej.

www.kwmimkm.polisi.pl

- Dlatego taki sam impuls na wejściu komórki może powodować inne jej pobudzenie niż dla innego wejścia.

STATYSTYKA:

- Liczba komórek nerwowych w mózgu: ok. 10^{11} ;
- Połączeń nerwowych – ok. 10^{14} ;
- Częstotliwość biologicznego neuronu – ok. kilkaset Hz.



17

ZALETY SSN:

www.kwmimkm.polisi.pl

- Nie wymagają programowania (tylko uczenie);
- Mają zdolność uogólniania;
- Są wysoce odporne na szумы i zniekształcenia sygnału;
- Pomagają wykrywać istotne powiązania pomiędzy danymi.

Stosuje się je gdy istnieje duża złożoność zagadnienia i trudno jest jednoznacznie określić formalne kryteria, dla stworzenia programu komputerowego.



18

PRZYKŁADOWE ZASTOSOWANIA:

Wojsko. Przetwarzanie sygnałów i obrazów (radar, sonar, termowizja) dla identyfikacji celu lub źródła sygnału, analiza obrazów satelitarnych, teledetekcja.

Telekomunikacja. Wytyczanie tras połączeń i sterowanie, kodowanie, filtracja, alokacja i korekcja kanałów, kompresja sygnałów i obrazów dla skrócenia czasu transmisji.

Przetwarzanie mowy. Rozpoznawanie mowy i mówców, synteza mowy.

Przemysł. Prognoza zużycia energii i surowców, testowanie nieinwazyjne, kontrola jakości wyrobów, sterowanie produkcją.

Robotyka. Pojazdy autonomiczne, sterowanie ramieniem robota.

Biologia i medycyna. Analiza obrazów i sygnałów, wspomaganie diagnozy.

Usługi. Automatyczne przetwarzanie dokumentów drukowanych i pisanych, identyfikacja podpisów, wykrywanie fałszerstw, rozpoznawanie osób (biometria), analiza danych marketingowych.

Finanse. Predykcja wskaźników giełdowych, kursów wymiany walut, analiza ryzyka.

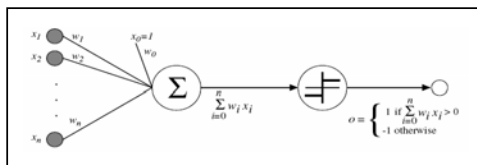
NIE NADAJĄ SIĘ DO:

- Przetwarzania informacji symbolicznej (np. edytory tekstu);
- Obliczeń o wymaganej wysokiej dokładności (sieć pracuje jakościowo, dając wyniki przybliżone);
- Rozwiązywania zagadnień, gdzie rozumowanie jest przeprowadzanie wieloetapowo (a musi być udokumentowane).



20

Perceptron (Rosenblatt 1958):



Układ posiadający wiele wejść i jedno wyjście:

Wejście: n stanów wejściowych x_1, \dots, x_n

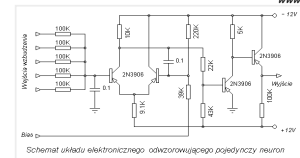
Wyjście: 0 (-1) lub 1

Uwaga: pod pojęciem „perceptronu” rozumie się też czasem sieć połączonych jednostek (neuronów).



21

Sprzętowo:



Pojedynczy perceptron pozwala na:

- przetwarzanie jednostkowych informacji;
- podejmowanie prostych decyzji;
- przekazywanie wyników sąsiadom.

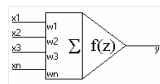
Dopiero w połączeniu z innymi węzłami uzyskuje się zdolność podejmowania złożonych decyzji.



22

Sygnał wyjściowy y_i i -tego neuronu liniowego:

$$y_i = \sum_{j=0}^N w_{ij} x_j$$



w_{ij} – waga dla j -tego wejścia i -tego neuronu;

x_j – j -ty sygnał wejściowy;

N – liczba wejść w i -tym neuronie.

Sygnał wyjściowy y_i i -tego neuronu (ogólnie):

$$y_i = \varphi(e) = \varphi \left(\sum_{j=1}^N w_{ij} x_j + B \right)$$

e – łączne pobudzenie neuronu (net value);

φ – funkcja aktywacji;

B – próg (bias).

W neuronie liniowym e jest sygnałem wyjściowym

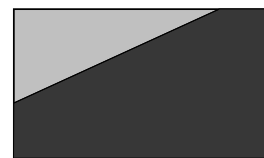


23

UCZENIE PERCEPTRONU:

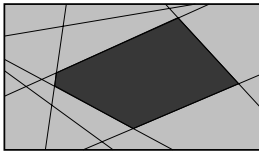
Pojedynczy neuron (lub 1 warstwa neuronów) typu perceptronowego jest w stanie rozdzielić przestrzeń obszarów wejściowych granicą decyzyjną na 2 obszary (półprzestrzenie).

- 2 zmienne – prosta;
- ogólnie – hiperpłaszczyzna.

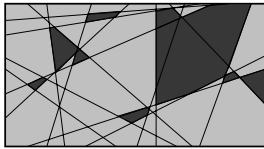


24

Sieć dwuwarstwowa – wielokąty wypukłe (simpleksy):

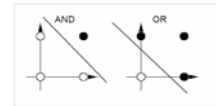
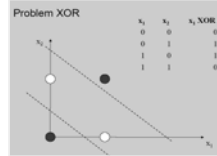
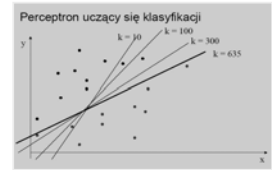
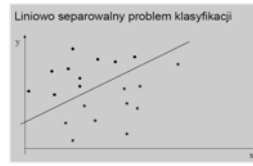


Sieci trzy- i więcej warstwowa – dowolne obszary (w tym wielokąty niewypukłe i obszary wielospójne).



25

Perceptron może prawidłowo klasyfikować sygnały, jeśli są liniowo separowalne:



26

Poprawka wartości wagi dla perceptronu w j -ym kroku (reguła delta):

$$\nabla w_i^{(j)} = \eta \delta^{(j)} x_i^{(j)}$$

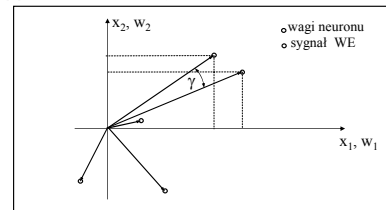
$$\delta^{(j)} = z^{(j)} - y^{(j)}$$

- z – wymagana odpowiedź neuronu;
- y – uzyskana odpowiedź neuronu;
- x – dana wejściowa dla i -tego wejścia;
- η – współczynnik uczenia (learning rate).



27

Sygnal WY jest tym większy, im bardziej sygnał WE będzie przypominał wagę (dla. sygnałów znormalizowanych).



Kąt pomiędzy wektorem WE a wektorem wag:

- mały – silny sygnał pozytywny;
- bliski 90° – słaby sygnał neutralny (bliski 0);
- duży – silny sygnał negatywny.

Dł. wektora WE znacząco mniejsza od dł. wektora wag – sygnał neutralny (niezależnie od kierunku wektora WE).



28

Waga:

- dodatnia - sygnał pobudzający;
- ujemna – sygnał gaszący;
- „0” - brak połączenia między neuronami.

Sygnały wyjściowe:

- $\langle 0,1 \rangle$ (funkcje unipolarne)
- $\langle -1,1 \rangle$ (funkcje bipolarne).



29

FUNKCJA AKTYWACJI

Wartość f . aktywacji – sygnał wyjściowy neuronu .

- liniowa f . aktywacji;
- nieliniowa f . aktywacji:

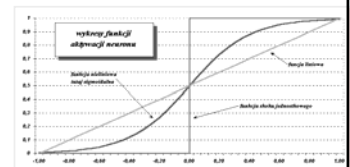
- f . skoku jednostkowego (progowa), np.:

$$\varphi(e) = \begin{cases} 1 & \text{gd}y \ e \geq 0 \\ 0 & \text{gd}y \ e < 0 \end{cases}$$

- inna, np. typu sigmoidalnego (f . logistyczna):

$$\varphi(e) = \frac{1}{1 + \exp(-\beta e)}$$

β – współczynnik sterujący nachyleniem krzywej



30

WYMAGANE CECHY F. AKTYWACJI:

- Ciągłe przejście pomiędzy wartością maksymalną a minimalną.
- Łatwa do obliczenia i ciągła pochodna

np. dla f. sigmoidalnej:

$$\varphi(e) = \frac{1}{1 + \exp(-\beta e)}$$

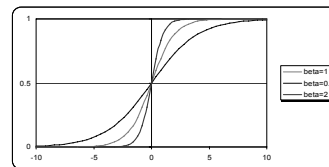
$$\varphi'(e) = \beta \cdot \varphi(e) \cdot [1 - \varphi(e)]$$

- Możliwość wprowadzenia do argumentu parametru β do ustalania kształtu krzywej.

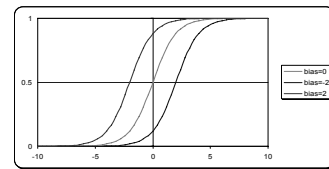


31

WSPÓŁCZYNNIK β



BIAS



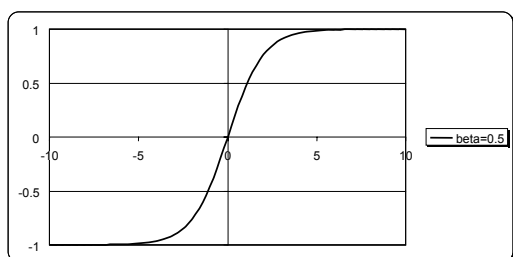
32

32

Bipolarny odpowiednik f. sigmoidalnej:

$$\varphi(e) = \frac{\exp(\beta e) - \exp(-\beta e)}{\exp(\beta e) + \exp(-\beta e)} = \text{tgh}(-\beta e)$$

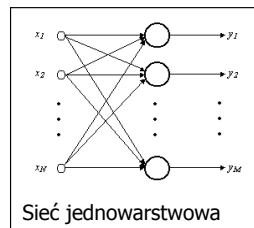
$$\varphi'(e) = \beta \cdot [1 + \varphi(e)] \cdot [1 - \varphi(e)]$$



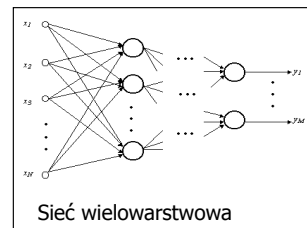
33

SIEĆ NEURONOWA:

- układ połączonych neuronów (model warstwowy)



Sieć jednowarstwowa



Sieć wielowarstwowa

Sieć wielowarstwowa – min. jedna warstwa ukryta.

Zwykle: kilka warstw, połączenia każdy z każdym (min. liczba parametrów do opisu).



34

RODZAJE NEURONÓW:

- **warstwy WE** (nie liczona);
- **warstw ukrytych:**
 - efekty działania obserwowane pośrednio poprzez WY;
 - pośredniczenie między WE a WY;
 - niemożność dokładnego obliczenia błędów;
- **warstwy WY** (rozwiązania stawianych zadań).

Pamięć neuronu – reprezentowana poprzez wagi.

Sieć działa jako **całość**.



35

KLASY SIECI:

- **feed-forward** (jednokierunkowe) – najczęściej stosowane;
- **rekurencyjne** – sprzężenia zwrotne;
- **sieci Kohonena** – samoorganizujące się;
- **sieci radialne** (RBF) - uniwersalny aproksymator.

STRUKTURA SIECI:

- ważna, lecz nie do przesady:

- wystarczający „potencjał intelektualny”;
- „rozsądna”.



36

Projektowanie sieci - problemy:

- ile warstw ukrytych?
- ile neuronów w warstwach?

Liczba neuronów w warstwie WE:

zależy od liczby danych podawanych na wejście.

Liczba neuronów w warstwie WY:

zależy od liczby poszukiwanych rozwiązań.

- Sieć z **1 warstwą ukrytą** powinna nauczyć się rozwiązywania większości postawionych problemów.
- Zbyt wiele warstw ukrytych – pogorszenie procesu uczenia (algorytm „grzeźnie w szczegółach”).



37

www.kwmimkn.polsl.pl

- Nieznane są problemy wymagające sieci z **więcej niż 3** warstwami ukrytymi (zwykle 1 lub 2).
- Liczbę neuronów w warstwie ukrytej można próbować oszacować:

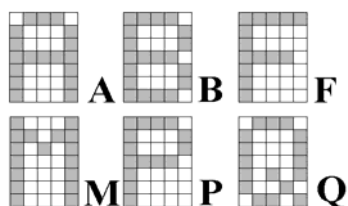
$$N_u = \sqrt{N_{we} \cdot N_{wy}}$$

- **Zwykle:** uczenie z początkowo małą liczbą neuronów i stopniowe zwiększanie ich liczby.
- Zbyt wiele elementów warstwy ukrytej: uczenie się „na pamięć”.



38

Przykład: Rozpoznawanie znaków alfabetu



- **WE - 35** (pikseli)
- **WY - 26** (liter)

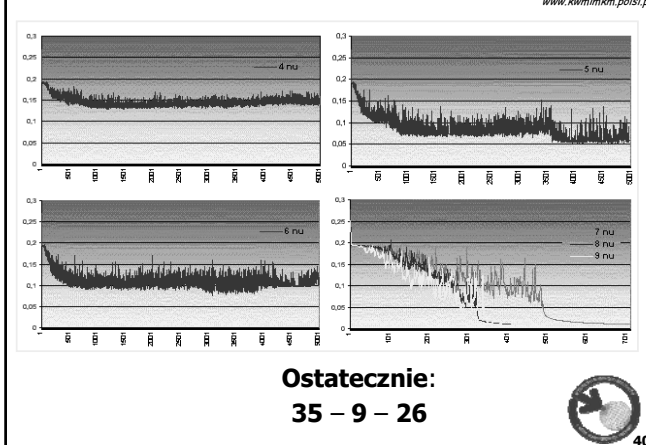
a co wewnątrz?

Wstępnie:
35 – 4 – 26



39

www.kwmimkn.polsl.pl



Ostatecznie:
35 – 9 – 26



40

UCZENIE SIECI NEURONOWYCH

Uczenie zamiast programowania.

Ta sama sieć może służyć do rozwiązywania skrajnie różnych zadań.

Uczenie sieci:

Wymuszanie określonego reagowania sieci na zadane sygnały wejściowe (poprzez odpowiedni dobór wag).

Uczenie sieci:

- uczenie **z nauczycielem** (nadzorowane);
- uczenie **z krytykiem**;
- uczenie **bez nauczyciela** (nienadzorowane).



41

www.kwmimkn.polsl.pl

UCZENIE Z NAUCZYCIELEM:

- Podawanie sieci zestawów sygnałów WE wraz z **prawidłowym** sygnałem WY.
- Naśladowanie **nauczyciela**, jakim jest **ciąg uczący**.
- Zestawy sygnałów (zwykle) powtarza się wielokrotnie, zaś sieć modyfikuje wagi na wejściach tak, by **zminimalizować błąd**.
- Zmiana wagi na i -tym wejściu neuronu po pokazaniu j -ego obiektu uczącego jest proporcjonalna do popełnianego na tym etapie błędu $\delta^{(j)}$.



42

UCZENIE SIECI NIELINIOWYCH

www.kwmimkm.polski.pl

Uczenie sieci – minimalizacja funkcji błędu.



- Zwykle gradientowe metody optymalizacji (np. metoda największego spadku).
- Warunek – funkcja aktywacji jest ciągła.



43

Poprawka wartości wagi:

www.kwmimkm.polski.pl

$$\nabla w_i^{(j)} = \eta \delta^{(j)} \frac{d\varphi(e)}{de^{(j)}} x_i^{(j)}$$

j – numer kroku uczenia;

- Jeśli neuron nie popełnia błędów wagi nie są zmieniane.
- Poprawka wagi na WE jest tym większa, im większy jest błąd na WY.
- Poprawka wagi na i -tym WE jest proporcjonalna do wielkości sygnału na tym wejściu (x_i).



44

Uaktualnianie wag:

www.kwmimkm.polski.pl

- po każdorazowej prezentacji jednej pary uczącej;
- po zaprezentowaniu wszystkich przykładów.

Przykładowy ciąg uczący (3 neurony, 5 wejść)

Typowy obiekt

który powinien być rozpoznawany przez **pierwszy** neuron:

WE: 3, 4, 3, 4, 5

WY: 1, -1, -1

Typowy obiekt

który powinien być rozpoznawany przez **drugi** neuron:

WE: 1, -2, 1, -2, -4

WY: -1, 1, -1

Typowy obiekt

który powinien być rozpoznawany przez **trzeci** neuron:

WE: -3, 2, -5, 3, 1

WY: -1, -1, 1



45

Nietypowy obiekt

który powinien być rozpoznawany przez **pierwszy** neuron:

WE: 4, 2, 5, 3, 2

WY: 0.8, -1, -1

Nietypowy obiekt

który powinien być rozpoznawany przez **drugi** neuron:

WE: 0, -1, 0, -3, -3

WY: -1, 0.8, -1

Nietypowy obiekt

który powinien być rozpoznawany przez **trzeci** neuron:

WE: -5, 1, -1, 4, 2

WY: -1, -1, 0.8

Obiekt

który powinien być odrzucany przez **wszystkie** neurony:

WE: -1, -1, -1, -1, -1

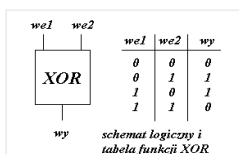
WY: -1, -1, -1



46

Funkcja logiczna XOR

www.kwmimkm.polski.pl



Schemat sieci:
2 - 2 - 1

Ciąg uczący:

WE 1	WE 2	WY
0	0	0
1	0	1
0	1	1
1	1	0

Ciąg weryfikujący (np):

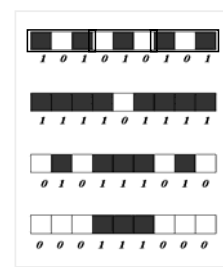
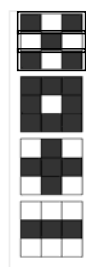
WE 1	WE 2
0.05	0.05
0.95	0.05
0.05	0.95
0.95	0.95



47

Rozpoznawanie znaków X, 0, +, -

www.kwmimkm.polski.pl



Matryca znaków

Liniowe rozwinięcie

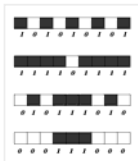
i zamiana na wektor WE

Schemat sieci: **9 - 5 - 4**



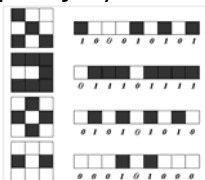
48

Ciąg uczący:



Wektor WE	Wektor WY
1 0 1 0 1 0 1 0 1	1 0 0 0
1 1 1 1 0 1 1 1 1	0 1 0 0
0 1 0 1 1 1 0 1 0	0 0 1 0
0 0 0 1 1 1 0 0 0	0 0 0 1

Nauczona sieć rozpoznaje symbole zniekształcone:



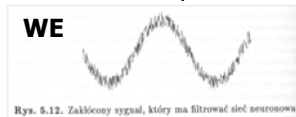
49

FILTROWANIE SYGNAŁÓW

WE - wzorce sygnałów zaszumionych;

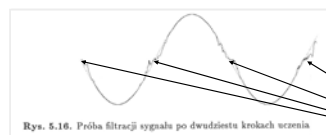
WY – sygnały czyste

np. sinus :



Rys. 5.12. Zakłócony sygnał, który ma filtrować sieć neuronowa

Rys. 5.13. Sygnał stworzony, który ma odtworzyć sieć neuronowa



Rys. 5.16. Próba filtracji sygnału po dwudziestu krokach uczenia



Z: R. Tadeusiewicz: „Elementarne wprowadzenie do techniki sieci neuronowych...”, PLJ, Warszawa, 1998

50

Okolice zera – mała wartość sygnału (trudności).

Rozwiązanie – przesunięcie sygnału:



Rys. 5.18. Próba filtracji sygnału przesuniętego po jednym kroku uczenia



Rys. 5.19. Próba filtracji sygnału przesuniętego po dwudziestu krokach uczenia



Z: R. Tadeusiewicz: „Elementarne wprowadzenie do techniki sieci neuronowych...”, PLJ, Warszawa, 1998

51

WSPÓŁCZYNNIK UCZENIA

η – learning rate

Wartości:

w przedziale 0.01÷5.0 (typowo **0.8÷2.0**)

- zbyt mały – powolne uczenie sieci;
- zbyt duży – gwałtowne zmiany parametrów sieci.

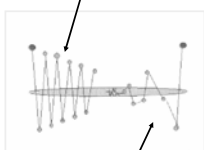


52

METODA MOMENTUM (ang. pęd)

- Nadanie uczeniu pewnej bezwładności;
- Zwiększenie szybkości uczenia bez zaburzenia stabilności algorytmu;

bez wsp. momentum



wsp. momentum = 0.5



53

Dodatkowy składnik:

zmiany wag zależą od błędów aktualnych i poprzednich.

$$\nabla w_i^{(j)} = \eta \delta^{(j)} \frac{d\varphi(e)}{de^{(j)}} x_i^{(j)} + \eta_2 \nabla w_i^{(j-1)}$$

η_2 – wartości w przedziale 0÷1 (często **0.9**).

Wagi: wartości początkowe zwykle losowo, często z zakresu (-0.1, 0.1) (najlepiej bez zera).



54