



# Sztuczne sieci neuronowe

Krzysztof A. Cyran  
POLITECHNIKA ŚLĄSKA  
Instytut Informatyki, p. 311

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- Problem inicjalizacji wag
- Problem doboru architektury sieci
- Dobór danych uczących
- Preprocessing: ekstrakcja cech

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- Problem inicjalizacji wag
- Problem doboru architektury sieci
- Dobór danych uczących
- Preprocessing: ekstrakcja cech

# Metody doboru współczynników uczenia

- Stałe współczynniki takie same dla wszystkich warstw (np.  $\eta = 0.05$ ,  $\mu = 0.9$ )
- Stałe współczynniki (osobno dla każdej z warstw)

$$\eta \leq \min(1/n_i)$$

gdzie:  $n_i$  to liczba wejść  $i$ -tego neuronu w rozważanej warstwie

- Cechy powyższych metod: proste ale mało wydajne

# Metody doboru współczynników uczenia (cd.)

- Metoda adaptacyjna (jeżeli błąd maleje to należy zwiększyć  $\eta$ ). Zdefiniujemy błąd  $\varepsilon$  jako:

$$\varepsilon = \sqrt{\sum_{j=1}^M (y_j - d_j)^2}$$

Jeżeli  $\varepsilon_i > k_w \varepsilon_{i-1}$  to:  $\eta_{i+1} = \eta_i \rho_d$  ( $\rho_d < 1$ )

w przeciwnym razie:  $\eta_{i+1} = \eta_i \rho_i$  ( $\rho_i > 1$ )

Wartości przykładowe:

współczynnik dopuszczalnego wzrostu błędu:  $k_w = 1.04$ ,

współczynnik zmniejszania  $\eta$ :  $\rho_d = 0.7$ ,

współczynnik zwiększania  $\eta$ :  $\rho_i = 1.05$ )

# Metody doboru współczynników uczenia (cd.)

Metoda doboru przez minimalizację kierunkową: jest to sposób najlepszy ale i najbardziej skomplikowany. Polega na znalezieniu minimum funkcji na wybranym kierunku (dowolną metodą) tzn. należy tak dobrać wartość współczynnika  $\eta$ , aby punkt  $\mathbf{W}_{k+1} = \mathbf{W}_k + \eta_k \mathbf{p}_k$  odpowiadał minimum funkcji  $E(\mathbf{W})$  na kierunku  $\mathbf{p}_k$ . Można stosować tu metody gradientowe, bezgradientowe (np. bisekcji) lub mieszane.

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- **Problem inicjalizacji wag**
- Problem doboru architektury sieci
- Dobór danych uczących
- Preprocessing: ekstrakcja cech

# Metody inicjalizacji wag

- Ponieważ od inicjalizacji wag zależy efekt procesu uczenia (minima lokalne!), zatem należy inicjalizować właściwie.
- Co to znaczy właściwie? – to znaczy w niecce przyciągania minimum globalnego
- Ze względu na to, że do końca nie da się określić jak to robić, zawsze należy stosować metodę multistartu i wybrać najlepszą sieć.



# Metody inicjalizacji wag (cd.)

- Metody losowej inicjalizacji (najczęściej stosowane ze względu na prostotę)
- Metody wykorzystujące uczenie nienadzorowane:
  - inicjalizacja z wykorzystaniem uogólnionej reguły Hebba
  - Inicjalizacja za pomocą algorytmu samoorganizacji

# Metody losowej inicjalizacji wag

Typowy przedział randomizacji:

$$\left[ -\frac{a}{\sqrt{n_{in}}}, \frac{a}{\sqrt{n_{in}}} \right]$$

dla funkcji sigmoidalnej:  $a = 2.38$ ,

$n_{in}$  – liczba wejść neuronu

# Metody losowej inicjalizacji wag (cd.)

Na podstawie badań statystycznych nad dynamiką szkła spinowego oraz analogii z procesem uczenia sieci neuronowych ustalono inny przedział randomizacji:

$$\begin{cases} \left[ -\frac{\sqrt{n_{in}}}{2}, \frac{\sqrt{n_{in}}}{2} \right] & \text{Dla neuronów warstw ukrytych} \\ 0, & \text{Dla neuronów wyjściowych} \end{cases}$$

# Metody losowej inicjalizacji wag (cd.)

Jeszcze inny proponowany przez Widrowa i Nguena przedział to:

$$\left\{ \begin{array}{l} \left[ -\sqrt{\frac{N_n}{n_{in}}}, \sqrt{\frac{N_n}{n_{in}}} \right] \text{ Dla neuronów warstw ukrytych} \\ [-0.5, 0.5] \text{ Dla neuronów wyjściowych} \end{array} \right.$$

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- Problem inicjalizacji wag
- **Problem doboru architektury sieci**
- Dobór danych uczących
- Preprocessing: ekstrakcja cech

# Dobór architektury sieci

- Dla realizacji odwzorowania nieciągłego potrzeba w ogólności dwóch warstw ukrytych
- Twierdzenie Kołmogorowa: Do dowolnej aproksymacji skalarnej funkcji ciągłej wystarczy jedna warstwa ukryta o rozmiarze  $2N+1$  (gdzie  $N$  to ilość wejść)
- Nie jest to jednak liczba optymalna: zwykle wystarcza dużo mniej neuronów np.:  $\sqrt{NM}$  (gdzie  $M$  to ilość wyjść sieci) lub jeszcze mniej.

# Podstawowy dylemat w doborze architektury sieci

- Za duża liczba wag (funkcja liczby neuronów ukrytych) skutkuje słabymi zdolnościami generalizacyjnymi sieci (jest za dużo stopni swobody w uczeniu)
- Za mała liczba wag skutkuje trudnościami w uczeniu (za mało stopni swobody)
- Teoretycznie postulowana liczebność wzorców uczących tak by była ona około 20 razy większa niż liczba modyfikowanych wag jest warunkiem trudnym do spełnienia w praktyce

# Metody rozwiązania dylematu

- Metody redukcji rozmiaru warstwy ukrytej (start od dużej sieci):
  - metody wrażliwościowe (np. OBD – optimal brain damage, OBS – optimal brain surgeon, usuwają wagę/neuron dla których funkcja celu ma najmniejszą wrażliwość)
  - Metody funkcji kary (dodatkowa kara za duże wartości wag powoduje dążenie do ich zmniejszania, by w końcu po osiągnięciu pewnego progu niektóre z nich wyeliminować)
- Dodawanie neuronów (start od małej sieci) np. algorytm kaskadowej korelacji Fahlmana)



# Inne metody zwiększania zdolności generalizacyjnych

- Sieci neuronowe z rozszerzeniami funkcyjnymi (sieci Pao).
- Motywacja: dla problemów źle uwarunkowanych ( $N \ll M$ ) można stosować dodatkowe wejścia w postaci funkcji wejść podstawowych (rzutowanie w przestrzeń o większej liczbie wymiarów poprawia uwarunkowanie i zwiększa zdolności generalizacyjne poprzez uproszczenie architektury (np. stosując wejścia  $x_1$ ,  $x_2$  oraz  $x_1x_2$  można rozwiązać problem XOR bez warstwy ukrytej)

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- Problem inicjalizacji wag
- Problem doboru architektury sieci
- **Dobór danych uczących**
- Preprocessing: ekstrakcja cech

# zwiększanie zdolności uogólniania poprzez dobór danych uczących



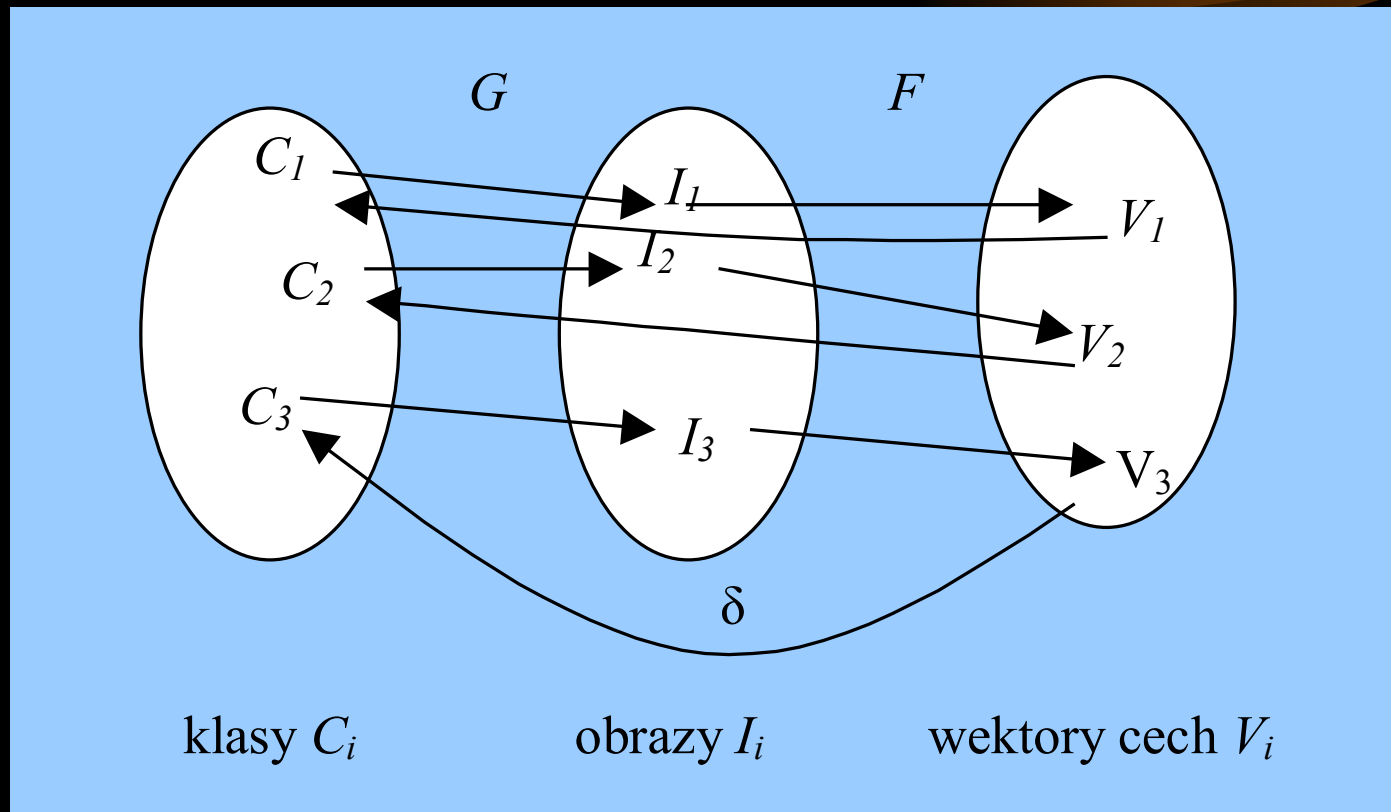
- Analiza wrażliwościowa danych uczących kończąca się wyborem tylko danych najistotniejszych
- Dodanie szumu do danych uczących

# Wykład 5

## PLAN:

- Metody doboru współczynnika uczenia
- Problem inicjalizacji wag
- Problem doboru architektury sieci
- Dobór danych uczących
- **Preprocessing: ekstrakcja cech**

# Odwzorowania w semantycznym rozpoznawaniu obrazów



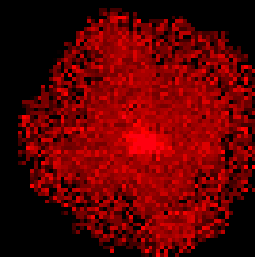
# Struktury plamkowe interferujących modów na wyjściu światłowodu



Jeden  
mod

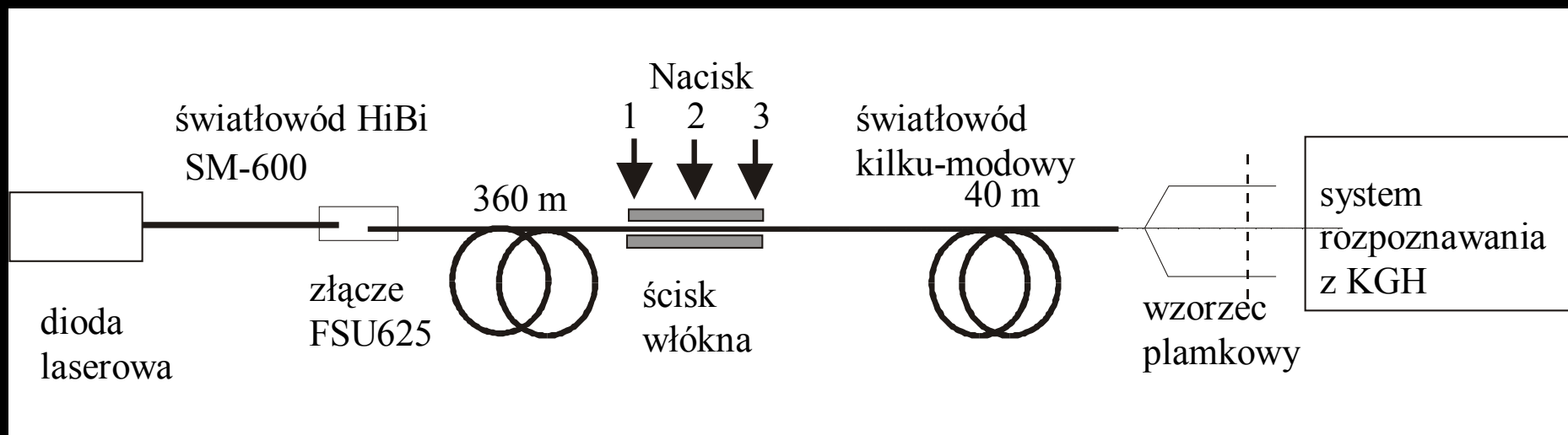


Kilka modów



Wiele  
modów

# Eksperymentalny układ symulujący naprężenia



# Optoelektroniczny system rozpoznawania obrazów z SSN jako klasyfikatorem

