Piotr Smuga gr2

Podstawy sztucznej inteligencji

Sprawozdanie nr 3

Cel: Celem ćwiczenia jest poznanie budowy i działania wielowarstwowych sieci neuronowych poprzez uczenie z użyciem algorytmu wstecznej propagacji błędu rozpoznawania konkretnych liter alfabetu.

Sieci jednokierunkowe których typowym przykładem jest perceptron jednowarstwowy (rys. perceptronu) składają się z neuronów ułożonych w warstwach o jednym kierunku przepływu sygnałów i połączeniach między warstwowych jedynie między kolejnymi warstwami. Sieć tego typu posiada warstwę wejściową, wyjściową i warstwy ukryte. Z funkcjonalnego punktu widzenia układ taki można traktować jako układ aproksymacji funkcji nieliniowej wielu zmiennych y = f(u).

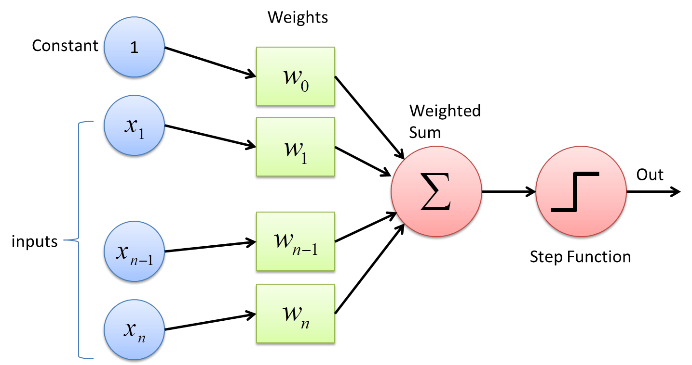


Figure 1 Perceptron

W moim projekcie użyłem sieci jednokierunkowej wielowarstwowej, zgodnie z założeniami.

Moja sieć zawiera 4 warstwy w tym 2 ukryte. Pierwsza warstwa posiada 25 perceptronów, druga 75 perceptronów, trzecia 50, natomiast ostatnia posiada 20 perceptronów

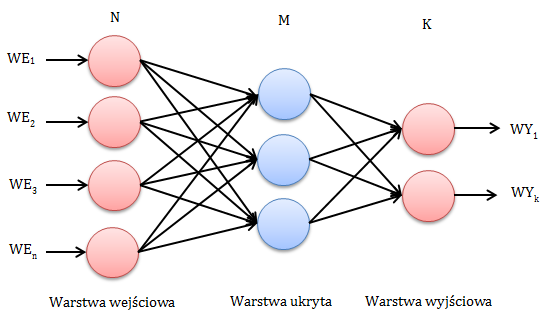


Figure 2 Przykład ilustrujący jednowarstwową jednokierunkową sieć neuronową

Sztuczny neuron to prosty system przetwarzający wartości sygnałów wprowadzanych na jego wejścia w pojedynczą wartość wyjściową, wysyłaną na jego jedynym wyjściu. To podstawowy element sieci neuronowych, która jest jedną z metod sztucznej inteligencji. Perceptron to prosty element obliczeniowy. Sumuje ważone sygnały wejściowe oraz porównuje ją z progiem aktywacji. Zależnie od wyniku perceptron może być wzbudzony albo nie. Do uczenia perceptronu można użyć algorytmu uczenia perceptronu, czyli automatycznego doboru wag na podstawie kolejnych przykładów.

Do uczenia perceptronu użyłem algorytmu RPROP (od ang. Resilient backPROPagation). Jest to algorytm przeznaczony dla pełnego (wsadowego) trybu korekcji parametrów (wag). Oznacza to, że jedno skorygowanie parametrów (wag) następuje dopiero po przeglądnięciu przez sieć całego zbioru uczącego i obliczeniu sumarycznego a tym samym dokładnego gradientu. Kluczowymi elementami algorytmu RPROP są: wykorzystywanie jedynie samego znaku każdej składowej gradientu (natomiast wartości są pomijane), a także modyfikowanie współczynnika (współczynników) uczenia w każdym kroku. Współczynnik uczenia jest zwiększany, gdy znaki kolejnych gradientów pozostają zgodne, natomiast zmniejszany (a dokładnie połowiony), gdy są różne.

Ogólny schemat procesu trenowania sieci wygląda następująco:

* Ustalamy topologię sieci, tzn. liczbę warstw, liczbę neuronów w warstwach.
* Inicjujemy wagi losowo (na małe wartości).
* Dla danego wektora uczącego obliczamy odpowiedź sieci (warstwa po warstwie).
* Każdy neuron wyjściowy oblicza swój błąd, oparty na różnicy pomiędzy obliczoną odpowiedzią y oraz poprawną odpowiedzią t.
* Błędy propagowane są do wcześniejszych warstw.
* Każdy neuron (również w warstwach ukrytych) modyfikuje wagi na podstawie wartości błędu i wielkosci przetwarzanych w tym kroku sygnałów.
* Powtarzamy od punktu 3. dla kolejnych wektorów uczących. Gdy wszystkie wektory zostaną użyte, losowo zmieniamy ich kolejność i zaczynamy wykorzystywać powtórnie.
* Zatrzymujemy się, gdy średni błąd na danych treningowych przestanie maleć. Możemy też co jakiś czas testować sieć na specjalnej puli nieużywanych do treningu próbek testowych i kończyć trenowanie, gdy błąd przestanie maleć.

- W programie użyłem funkcję Ramp jako funkcję aktywacji

- każda litera reprezentowana jest jako tablica zawierająca 5x5 liczb reprezentującą piksele

- dane uczące składają się z tablicy 20 liczb reprezentujących 20 pierwszych wielkich liter alfabetu angielskiego:

new double[25] {1,1,1,1,1,1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1};

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

new double[25] {1,1,1,0,0,1,0,0,1,0,1,1,1,0,0,1,0,0,1,0,1,1,1,0,0},

new double[25] {1,1,1,1,1,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,1,1,1,1},

new double[25] {1,1,1,0,0,1,0,1,0,0,1,0,1,0,0,1,0,1,0,0,1,1,1,0,0},

new double[25] {1,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,0,0,0,0,1,1,1,1,1},

new double[25] {1,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,0,0,0,0,1,0,0,0,0},

new double[25] {1,1,1,1,0,1,0,0,0,0,1,0,1,1,0,1,0,0,1,0,1,1,1,1,0},

new double[25] {1,0,0,0,1,1,0,0,0,1,1,1,1,1,1,1,0,0,0,1,1,0,0,0,1},

new double[25] {1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0},

new double[25] {1,1,1,0,0,0,0,1,0,0,0,0,1,0,0,1,0,1,0,0,1,1,1,0,0},

new double[25] {1,0,0,1,0,1,0,1,0,0,1,1,0,0,0,1,0,1,0,0,1,0,0,1,0},

new double[25] {1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,1,1,0,0},

new double[25] {1,1,0,1,1,1,0,1,0,1,1,0,0,0,1,1,0,0,0,1,1,0,0,0,1},

new double[25] {1,1,0,0,1,1,1,1,0,1,1,0,1,1,1,1,0,0,1,1,1,0,0,0,1},

new double[25] {1,1,1,1,1,1,0,0,0,1,1,0,0,0,1,1,0,0,0,1,1,1,1,1,1},

new double[25] {1,1,1,1,0,1,0,0,1,0,1,1,1,1,0,1,0,0,0,0,1,0,0,0,0},

new double[25] {1,1,1,1,0,1,0,0,1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,0},

new double[25] {1,1,1,0,0,1,0,0,0,0,1,1,1,0,0,0,0,1,0,0,1,1,1,0,0},

new double[25] {1,1,1,1,1,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0},

new double[25] {1,0,0,1,0,1,0,0,1,0,1,0,0,1,0,1,0,0,1,0,1,1,1,1,0}

- dane testujące składają się z tablicy 25 liczb reprezentujących daną literkę: new double[25] { 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0 }

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Wyniki:

Wyniki uczenia:

A: 0,632 0,000 0,000 0,000 0,000 0,291 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

B: 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

C: 0,000 0,000 0,983 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,010 0,000 0,000 0,000 0,000 0,000

D: 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,030 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

E: 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

F: 0,356 0,000 0,000 0,000 0,000 0,663 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

G: 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

H: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,917 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

I: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,978 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

J: 0,000 0,000 0,000 0,000 0,000 0,062 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

K: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

L: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,952 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

M: 0,000 0,000 0,000 0,000 0,036 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000

N: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000 0,000 0,000

O: 0,000 0,000 0,017 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,942 0,004 0,000 0,000 0,000 0,000

P: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000 0,000

R: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000 0,000

S: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000 0,000

T: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000 0,000

U: 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 0,000 1,000

Wnioski:

* Im mniejszy współczynnik uczenia tym dokładniejszy wynik. Aby otrzymać wiarygodne dane musimy odpowiednio dobrać współczynniki uczenia a także bezwładności
* wraz ze zwiększaniem wartości współczynnika uczenia zmniejsza się liczba iteracji, co ma wpływ na działanie programu i powoduje uzyskanie błędnych wyników. Aby zredukować liczbę błędnych wyników można zwiększyć liczbę iteracji.
* dobór wag ma bezpośredni wpływ na działanie perceptronów w sieci, w zależności od ich wartości zmienia się poprawność wyników i liczba iteracji. Wartości wag mają największy wpływ na efekt końcowy, często wagi są ustalane losowo.
* dane uczące wpływają na poprawność uczenia perceptronu, przy zbyt małej ilości otrzymujemy błędne wyniki. Aby uzyskać lepsze wyniki powinniśmy dostarczyć wystarczającą liczbę danych wejściowych
* bardzo ważny jest odpowiedni dobór współczynnika uczenia, wag oraz liczby danych uczących, gdyż mają bezpośredni wpływ na działanie sieci.
* Zastosowanie sieci wielowarstwowych umożliwia uzyskanie dokładniejszych wyników w krótszym czasie obliczeń.

Listing kodu:

* Cały listing kodu został umieszony w repozytorium Git pod następującym adresem: https://github.com/psmuga/PSI/blob/master/Scenariusz%203/ConsoleApp1/ConsoleApp1/Program.cs