

# Dokumentacja implementacji algorytmu wstecznej propagacji na przykładzie funkcji NXoR

Michał Korniak

## Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>2</b>
<b>2</b>	<b>Algorytm wstecznej propagacji</b>	<b>2</b>
2.1	Sposób działania wielowarstwowych sieci neuronowych . . . . .	2
2.2	Implementacja wielowarstwowej sieci neuronowej . . . . .	4
2.3	Trening sieci neuronowej . . . . .	6

# 1 Wprowadzenie

Algorytm wstecznej propagacji błędów jest metodą uczenia wielowarstwowych polegającą na korekcie wag połączeń między neuronami na podstawie błędów całej sieci. W tym dokumencie prześledzę działanie tego algorytmu, jednocześnie przedstawiając jego autorską implementację w języku C#. Następnie pokażę jak algorytm radzi sobie z nauczaniem funkcji NXoR.

## 2 Algorytm wstecznej propagacji

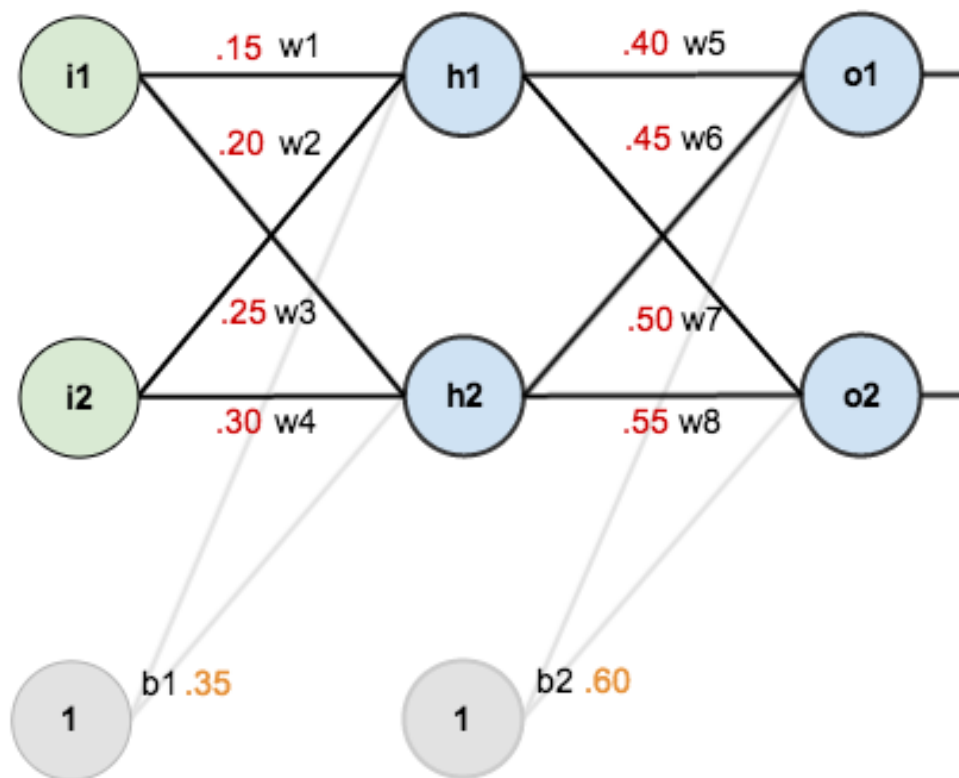
### 2.1 Sposób działania wielowarstwowych sieci neuronowych

Algorytm wstecznej propagacji jest wykorzystywany do nauczania wielowarstwowych sieci jednokierunkowych. Taka sieć zawiera warstwę wejściową i warstwę wyjściową, może posiadać również warstwy ukryte (w zależności od problemu jedną albo dwie). Każda warstwa zawiera dowolną ilość neuronów. Neurony są połączone ze sobą w taki sposób, że każdy neuron warstwy innej niż wyjściowa jest połączony z każdym neuronem kolejnej warstwy, a każde połączenie ma określoną wagę. Dodatkowo możliwe jest połączenie do tego zwanego biasa, czyli połączenia do neuronu, który zawsze przyjmuje wartość 1. Połączenia wejściowe do neuronu będą wpływać na to jaką będzie miał wartość.

Rysunek 1 przedstawia przykład jednokierunkowej sieci neuronowej opartej o dwa neurony warstwy wejściowej i tyle samo neuronów w warstwach ukrytej i wyjściowej. Oprócz połączeń do innych neuronów, istnieją również połączenia do biasa, który jest wspólny dla neuronów w ramach jednej warstwy.

Rysunek 1 zawiera wartości neuronów wejściowych oraz wag, dzięki czemu będziemy mogli pokazać jak wyliczane są wartości kolejnych neuronów. Robimy to sumując wartości połączeń do tego neuronu, a następnie poddając wynik funkcji aktywacji. Przez wartość połączenia będziemy rozumieli iloczyn wagi i neuronu wejściowego dla połączenia. Przykładowo dla neuronu  $h1$  wartość przed użyciem funkcji aktywacji będzie równa:

$$net_{h1} = i1 * w1 + i2 * w2 + 1 * b1$$



Źródło: <https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

Rysunek 1: Przykład wielowarstwowej sieci jednokierunkowej

Co po podstawieniu wartości da:

$$net_{h1} = 0.05 * 0.15 + 0.10 * 0.20 + 1 * 0.35 = 0.3775$$

Taka wartość jest następnie poddawana działaniu funkcji aktywacji. Takie działanie umożliwia normalizowanie wartości neuronów do oczekiwanych przedziałów wartości. W tym przypadku skorzystamy z funkcji sigmoidalnej, której zakres wartości mieści się w przedziale  $[0,1]$ :

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

Po podstawieniu otrzymanej wcześniej wartości do funkcji dostaniemy wartość neuronu  $h_1$ :

$$out_{h_1} = f(net_{h_1}) = \frac{1}{1 + e^{-net_{h_1}}} = \frac{1}{1 + e^{-0.3775}} \approx 0.5932$$

W ten sposób będziemy liczyć wartość kolejnych neuronów aż otrzymamy wynik.

## 2.2 Implementacja wielowarstwowej sieci neuronowej

Teraz przedstawię implementację mechanizmu zaprezentowanego w poprzednim podrozdziale. Jej podstawą jest klasa odwzorowująca sieć neuronową, w naszym przypadku taką funkcję pełni klasa `NeuralNetwork`, która to implementuje interfejs pokazany na listingu 1.

Listing 1: Interfejs `INeuralNetwork`

```
1 public interface INeuralNetwork
2 {
3     IEnumerable<InputNeuron> InputLayer { get; }
4     IEnumerable<IHiddenNeuron> HiddenLayer { get; }
5     IEnumerable<IOutputNeuron> OutputLayer { get; }
6     IErrorFunction ErrorFunction { get; }
7     void FillInputNeurons(IEnumerable<double> input);
8     IEnumerable<double> CalculateOutput();
9 }
```

W tym momencie skupmy się na liniijkach 3, 4 i 5, które to wskazują na to, że sieć neuronowa zawiera warstwy wejściową, ukrytą i wyjściową. Jak widzimy każda warstwa posiada inny typ neuronu, co wynika to z tego, że w zależności od warstwy neurony się różnią. Przykładowo:

- Neuron warstwy wejściowej nie ma możliwości dodawanie połączeń wejściowych.
- Neuron warstwy wyjściowej nie ma możliwości dodawanie połączeń wyjściowych.
- Neuron warstwy ukrytej ma możliwość dodawanie obu typów połączeń.

Obiekt klasy `NeuralNetwork` jest tworzony z wykorzystaniem wzorca "Builder". Dzieje się to w sposób pokazany na listingu 2.

Listing 2: Budowanie obiektu NeuralNetwork

```

1  var neuralNetworkBuilder = new NeuralNetworkBuilder();
2  var network = neuralNetworkBuilder
3      .SetNumberOfInputNeurons(2)
4      .SetNumberOfOutputNeurons(1)
5      .SetActivationFunction(new SigmoidActivationFunction())
6      .SetErrorFunction(new MeanSquaredErrorFunction(1))
7      .SetNumberOfHiddenNeurons(3) // opcjonalne
8      .AddBiasConnections()        // opcjonalne
9      .Build();

```

Klasa NeuralNetworkBuilder umożliwia stworzenie sieci neuronowej:

- Zawierającą wybraną ilość neuronów warstwy wejściowej.
- Zawierającą wybraną ilość neuronów warstwy wyjściowej
- Działającą na określonej funkcji aktywacji
- Wyliczającą błąd na podstawie wybranej funkcji
- Umożliwiającej opcjonalne dodanie warstwy ukrytej
- Umożliwiającej opcjonalne dodanie połączeń do biasów

Zadaniem Buildera jest stworzenie obiektu klasy NeuralNetwork, który spełni podane wymagania. W zależności od tego czy użytkownik będzie potrzebował warstwy ukrytej, Builder utworzy sieć która ma połączenia z warstwą pośrednią lub też bezpośrednie połączenia warstwy wejściowej z wyjściową. Podobnie jest z wyborem tego czy powinny być tworzone biasy czy też nie. Builder nie posiada żadnej metody dotyczącej wyboru pierwotnych wag połączeń, co wynika z tego, że zgodnie z założeniem algorytmu wstecznej propagacji są one inicjalizowane losowo.

Dla tak stworzonej sieci neuronowej możemy przeprowadzić obliczenia, które zostały przedstawione w poprzednim podrozdziale. Robimy to w sposób przedstawiony w listingu 3

Listing 3: Budowanie obiektu NeuralNetwork

```

1  network.FillInputNeurons(new double[] { 0, 1 });
2  var output = network.CalculateOutput();

```

W pierwszej kolejności są wypełniane neurony wejściowe. Jeśli użytkownik poda liczbę danych wejściowych różniącą się od liczby neuronów warstwy wejściowej zostanie wyrzucony wyjątek.

Wyjście sieci jest liczone w sposób rekurencyjny:

- Wyjście każdego neuronu z wyjątkiem neuronów wejściowych jest sumą wyjść połączeń wejściowych poddaną funkcji aktywacji (listing 4).
- Wyjście połączenia jest iloczynem wagi i wyjścia neuronu źródłowego (listing 5)

Listing 4: Wylizanie wyjścia neuronu

```

1 public class OutputNeuron : IOutputNeuron
2 {
3     /**
4     public double NetOutput => _inputConnections.Sum(x => x.Output);
5     public double Output => _activationFunction.Invoke(NetOutput);
6     /**
7 }

```

Listing 5: Wylizanie wyjścia połączenia

```

1 class NeuronConnection : INeuronConnection
2 {
3     /**
4     public double Output => Weight * _source.Output;
5     /**
6 }

```

## 2.3 Trening sieci neuronowej

Wagi połączeń są inicjowane wartościami losowymi, w związku z tym możemy się spodziewać, że pierwotne wyjście sieci będzie znacznie odbiegało od tego czego oczekujemy. Dlatego też przeprowadzany jest trening sieci, którego celem jest ustawienie wag w ten sposób, żeby wyjście sieci jak najbardziej odpowiadało wartościom oczekiwanym.

Pierwszym krokiem jest podanie sieci danych testowych, które posiadają informację na temat wejścia oraz oczekiwanego wyjścia sieci. Następnie obliczane jest rzeczywiste wyjście neuronu, które są porównywane z oczekiwanym za pomocą funkcji błędu średniokwadratowego.

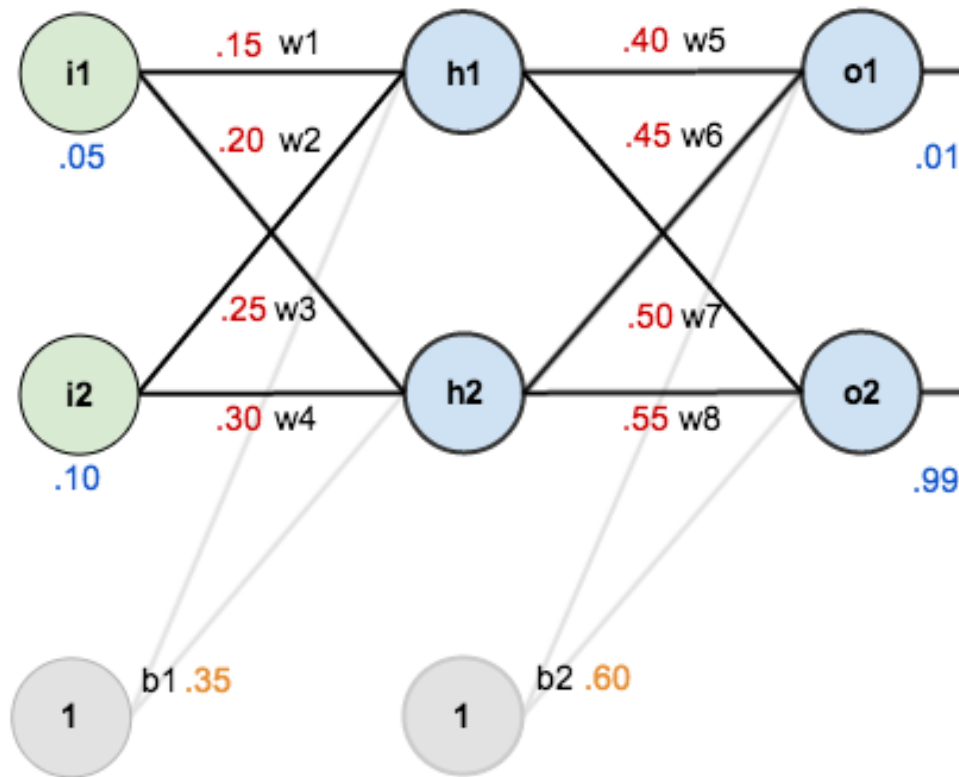
$$E(O) = 1/n * (d - y)^2$$

gdzie:

- $E(O)$  to błąd średniokwadratowy dla konkretnego neuronu wyjściowego
- $n$  to liczba neuronów warstwy wyjściowej

- d to oczekiwane wyjście neuronu
- y to rzeczywiste wyjście neuronu

Suma błędów średniokwadratowych wszystkich neuronów wyjściowych jest błędem całkowitym.



Źródło: <https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

Rysunek 2: Przykład wielowarstwowej sieci jednokierunkowej z podanymi wejściami oraz oczekiwanymi wyjściami

Rozważmy sieć z rysunku 2, na którym zostały podane wejścia oraz oczekiwane wyjścia sieci. Wartości neuronów warstwy wyjściowej będą równe:

$$out_{o1} \approx 0.7513$$

$$out_{o2} \approx 0.7729$$

Mając podane oczekiwane oraz rzeczywiste wyjście możemy obliczyć błąd średniokwadratowy dla obu neuronów:

$$E_{o1} = 1/2 * (0.01 - 0.7513)^2 \approx 0.274762845$$

$$E_{o2} = 1/2 * (0.99 - 0.7729)^2 \approx 0.023566205$$

Tym samym błąd całkowity wynosi:

$$E_{total} = E_{o1} + E_{o2} \approx 0.274762845 + 0.023566205 \approx 0.29832905$$

Jak widzimy sieć jest daleka do ideału, dlatego będziemy chcieli zmodyfikować wagi tak, aby jak najbardziej zminimalizować błąd. Robimy to za pomocą wzoru:

$$w_5^+ = w_5 - \eta * \frac{\partial E_{total}}{\partial w_5}$$

gdzie:

$w_5^+$  jest nową wartością wagi.

$w_5$  jest aktualną wartością wagi.

$\eta$  jest stałą uczenia, czyli wartością, która określa szybkość uczenia sieci.

$\frac{\partial E_{total}}{\partial w_5}$  jest pochodną częściową błędu całkowitego sieci do danej wagi.

Szczególną uwagę należy poświęcić stałej uczenia  $\eta$ . Jest ona ustawiana w momencie rozpoczynania procesu, przyjmując wartości z zakresu  $[0,1]$ . W zależności od wybranej wartości, zmiany wag będą postępować gwałtownie lub też łagodnie. Zazwyczaj stała uczenia jest ustawiana na wartości mniejsze od 0.1. Jednakże w tym przykładzie ustawimy ją na 0.5.

Kolejną rzeczą wartą uwagi jest pochodna błędu całkowitego do konkretnej wagi. Liczymy ją z wykorzystaniem reguły łańcuchowej, w przypadku  $w_5$ :

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5}$$

Już teraz możemy zauważyć, że wzór na pochodną cząstkową dla  $w_6$  będzie bardzo podobny:

$$\frac{\partial E_{total}}{\partial w_6} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_6}$$



Wynika to z tego, że oba połączenie mają taki neuron docelowy. Aby nie liczyć tego samego dwa razy, możemy wprowadzić oznaczenie dla części wspólnej równania:

$$\delta_{o1} = \frac{\partial E_{total}}{\partial net_{o1}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}}$$

tym samym:

$$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} * \frac{\partial net_{o1}}{\partial w_5}$$

$$\frac{\partial E_{total}}{\partial w_6} = \delta_{o1} * \frac{\partial net_{o1}}{\partial w_6}$$

Spróbujmy teraz uprościć wyrażenie  $\frac{\partial net_{o1}}{\partial w_5}$ :

$$\frac{\partial net_{o1}}{\partial w_5} = \frac{\partial(w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1)}{\partial w_5} = 1 * out_{h1} * w_5^{(1-1)} + 0 + 0 = out_{h1}$$

Jak widzimy  $\frac{\partial net_{o1}}{\partial w_5}$  ostatecznie sprowadza się do wartości neuronu źródłowego dla połączenia. W analogiczny sposób upraszcza się równanie dla  $\frac{\partial net_{o1}}{\partial w_6}$ , dzięki czemu otrzymujemy:

$$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} * out_{h1}$$

$$\frac{\partial E_{total}}{\partial w_6} = \delta_{o1} * out_{h2}$$

Ostatnim krokiem potrzebnym do wyliczenia nowych wag  $w_5$  jest obliczenie samego współczynnika delty dla  $o_1$  ( $\delta_{o1}$ ):

$$\delta_{o1} = \frac{\partial E_{total}}{\partial net_{o1}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}}$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = \frac{\partial(\frac{1}{2}(d_{o1} - out_{o1})^2 + \frac{1}{2}(d_{o2} - out_{o2})^2)}{\partial out_{o1}} \quad (1)$$

$$= 2 * \frac{1}{2}(target_{o1} - out_{o1})^{2-1} * -1 + 0 \quad (2)$$

$$= -(target_{o1} - out_{o1}) \quad (3)$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = \frac{\partial(\frac{1}{1+e^{-net_{o1}}})}{\partial net_{o1}} \quad (1)$$

$$= (\frac{1}{1+e^{-net_{o1}}})(1 - \frac{1}{1+e^{-net_{o1}}}) \quad (2)$$

gdzie:

$$\frac{1}{1+e^{-net_{o1}}} = out_{o1} \quad (3)$$

a więc:

$$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1}(1 - out_{o1}) \quad (4)$$

W związku z powyższymi równaniami wzór na współczynnik delty dla  $out_{o1}$  wynosi:

$$\delta_{o1} = \frac{\partial E_{total}}{\partial net_{o1}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} \quad (1)$$

$$= -(target_{o1} - out_{o1}) * out_{o1}(1 - out_{o1}) \quad (2)$$

$$(3)$$

$$\delta_{o1} = \frac{\partial E_{total}}{\partial net_{o1}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} \quad (1)$$

$$= -(target_{o1} - out_{o1}) * out_{o1}(1 - out_{o1}) \quad (2)$$

$$(3)$$

Co podstawieniu daje:

$$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} * out_{h1} = -(target_{o1} - out_{o1}) * out_{o1}(1 - out_{o1}) \approx 0.1385$$

W ten sposób otrzymaliśmy łatwy wzór na obliczenie pochodnych cząstkowej po  $w_5$  i  $w_6$

$$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} * out_{h1} \approx 0.1385 * 0.593269992 \approx 0.082167$$

$$\frac{\partial E_{total}}{\partial w_6} = \delta_{o1} * out_{h2} \approx 0.1385 * 0.596884378 \approx 0.082668$$

co umożliwia nam obliczenie nowych wag:

$$w_5^+ = w_5 - \eta * \frac{\partial E_{total}}{\partial w_5} = 0.4 - 0.5 * 0.082167 = 0.359$$

$$w_6^+ = w_6 - \eta * \frac{\partial E_{total}}{\partial w_6} = 0.45 - 0.5 * 0.082668 = 0.409$$

W analogiczny sposób liczymy nowe wagi połączeń z neuronem  $o_2$ :

$$w_7^+ = w_7 - \eta * \frac{\partial E_{total}}{\partial w_7} = w_7 - \eta * \delta_{o2} * out_{h1} = 0.511$$

$$w_8^+ = w_8 - \eta * \frac{\partial E_{total}}{\partial w_8} = w_8 - \eta * \delta_{o2} * out_{h2} = 0.561$$

Bias jest trochę inny w przypadku, gdyż jest połączony jednocześnie z neuronami  $o_1$  i  $o_2$ , w związku z tym pochodna cząstkowa do tej wagi będzie wyliczana w następujący sposób:

$$\frac{\partial E_{total}}{\partial b_2} = \frac{\partial E_1 + E_2}{\partial b_2} = \frac{\partial E_1}{\partial b_2} + \frac{\partial E_2}{\partial b_2} \quad (1)$$

$$= \left( \frac{\partial E_1}{\partial b_2} \right) \quad (2)$$

$$(3)$$

$$b_2^+ = b_2 - \eta * \frac{\partial E_{total}}{\partial b_2} = b_2 - \eta * \delta_{o1} + \delta_{o2} * out_{h2} = 0.561370121$$