

Dokumentacja implementacji algorytmu wstecznej propagacji na przykładzie funkcji NXoR

Michał Korniak

Spis treści

1	Wprowadzenie	2
2	Architektura sieci dla funkcji NXoR	2
3	Obliczanie wyjść sieci neuronowej	4
4	Trening sieci neuronowej	5
5	Implementacja wielowarstwowej sieci neuronowej	10

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 na przykładzie nauczania działania funkcji NXoR. jednocześnie przedstawiając jego autorską implementację w języku C#.

2 Architektura sieci dla funkcji NXoR

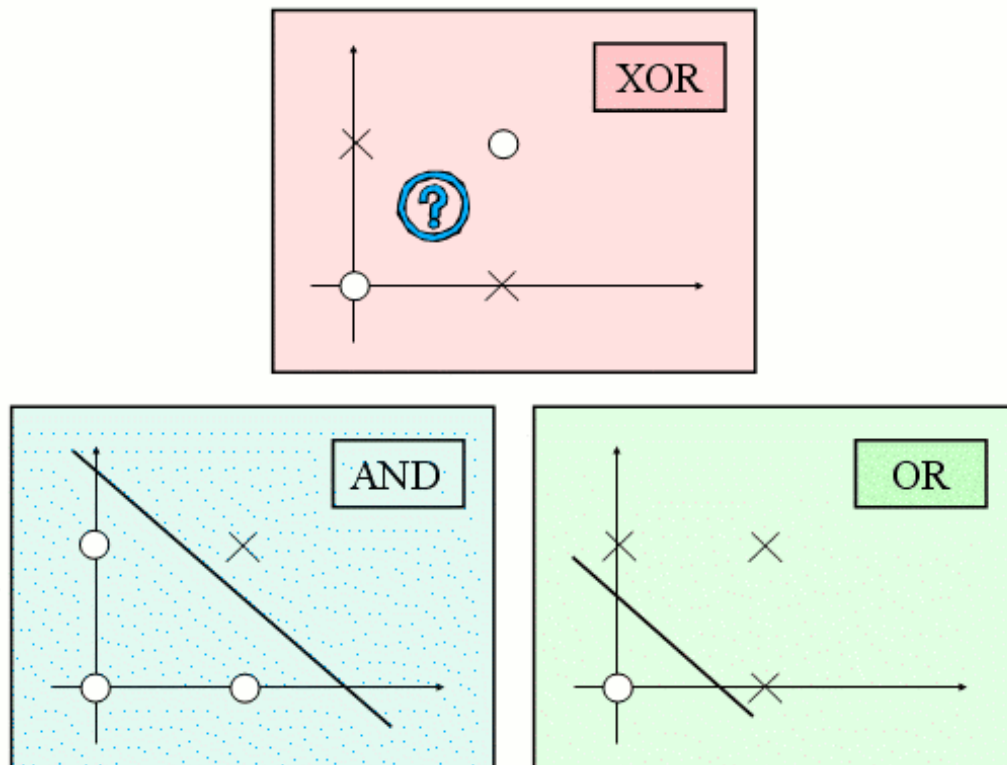
Algorytm wstecznej propagacji jest wykorzystywany do nauczania wielowarstwowych sieci jednokierunkowych. Taka sieć zawiera warstwę wejściową i warstwę wyjściową, może posiadać również warstwy ukryte (wykorzystywana do problemów liniowo nieseparowalnych). Każda warstwa zawiera dowolną ilość neuronów. 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.

Kluczowym pytaniem jest to jak powinna wyglądać sieć obsługująca funkcję NXoR. Jako, że funkcja NXoR przyjmuje dwie liczby wejściowe jasne jest, że powinna posiadać również dwa neurony wejściowe. Wiemy również, że sieć posiada jedno wyjście, co sprawia, że potrzebujemy tylko jednego neuronu wyjściowego.

x1	x2	NXoR
0	0	1
0	1	0
1	0	0
1	1	1

W powyższym opisie pojawiła się wzmianka o tym, że warstwa ukryta jest wykorzystywana w problemach, które nie są liniowo separowalne. W związku z tym należy się zastanowić czy problem NXoR do takowych należy. Aby sprawdzić czy problem jest liniowo separowalny należy narysować na płaszczyźnie punkty w których funkcja przyjmuje 0 i tych w których przyjmuje 1, a następnie sprawdzić czy istnieje prosta, która przedzieli te punkty

na dwie płaszczyzny. Jak pokazuje rysunek 1 do funkcji spełniających ten warunek należą funkcje AND i OR. Natomiast przykładem funkcji niespełniającej tego warunku jest XOR, a więc też jego odwrotność czyli NXoR. W związku z tym będziemy potrzebować warstwy ukrytej.



Źródło: <https://edux.pjwstk.edu.pl/mat/2144/lec/rW2.htm/>

Rysunek 1: Liniowa separowalność funkcji AND, OR, XOR

Otwartym pytaniem pozostaje natomiast to ile neuronów powinna posiadać ta warstwa. Tutaj posłużymy się cytatem z książki Jeffa Heaton "Introduction to Neural Networks for C#":

There are many rule-of-thumb methods for determining the correct number of neurons to use in the hidden layers, such as the following:

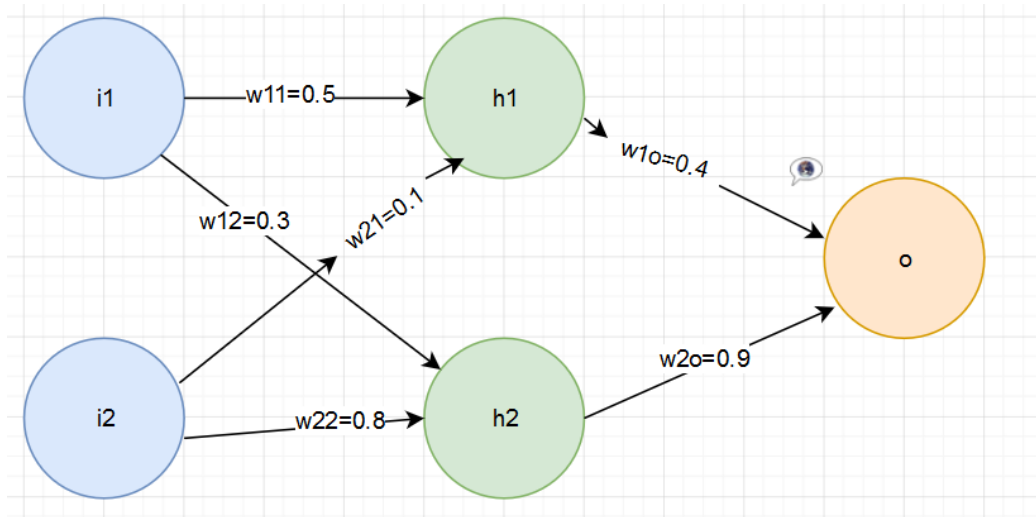
- The number of hidden neurons should be between the size of the input layer and the size of the output layer.

- The number of hidden neurons should be $2/3$ the size of the input layer, plus the size of the output layer.
- The number of hidden neurons should be less than twice the size of the input layer.

W związku z powyższym nasza warstwa ukryta będzie zawierała 2 neurony.

Kolejnym pytaniem jest to czy należy wprowadzać biasa. Wstępne testy pokazały, że jego obecność przyspiesza trening sieci, ale aby nie komplikować opisu zostanie on pominięty.

Ostatecznie nasza sieć będzie wyglądała tak jak na rysunku 2



Rysunek 2: Architektura sieci realizującej funkcję NXoR

3 Obliczanie wyjść sieci neuronowej

Nieodzownym krokiem treningu sieci neuronowej jest obliczenie aktualnych wyników i porównanie ich z oczekiwanymi. Przedstawmy to na przykładzie wejścia (0,1), którego oczekiwanym wyjściem jest 0. Informację o wejściu podajemy neuronom wejściowym, tak więc: $i1 = 0$, $i2 = 1$.

Aby obliczyć wartości kolejnych neuronów musimy zsumować wartości połączeń do tego neuronu, a następnie poddając wynik funkcji aktywacji. Przez

wartość połączenia będziemy rozumieli iloczyn wagi i neuronu wejściowego dla połączenia. Wagi połączeń zostały przedstawione na rysunku 1.

Przykładowo dla neuronu $h1$ wartość przed użyciem funkcji aktywacji będzie równa:

$$net_{h1} = i1 * w11 + i2 * w21$$

Co po podstawieniu wartości da:

$$net_{h1} = 0 * 0.5 + 1 * 0.1 = 0.1$$

Taka wartość jest następnie poddawana działaniu funkcji aktywacji, co 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 $h1$:

$$out_{h1} = f(net_{h1}) = \frac{1}{1 + e^{-net_{h1}}} = \frac{1}{1 + e^{-0.1}} \approx 0.525$$

W ten sposób będziemy liczyć wartość kolejnych neuronów:

$$out_{h2} = f(i1 * w12 + i2 * w22) = f(0 * 0.3 + 1 * 0.8) = f(0.8) \approx 0.690$$

$$out_o = f(h1 * w1o + h2 * w2o) = f(0.525 * 0.4 + 0.690 * 0.9) = f(0.831) \approx 0.697$$

4 Trening sieci neuronowej

Jak widzimy otrzymany wynik różni się od oczekiwanego. Dlatego też przeprowadzimy trening sieci, którego celem będzie ustawienie wag w ten sposób, żeby wyjście sieci jak najbardziej odpowiadało wartości oczekiwanej.

Mając obliczone wyjście sieci oraz wartością oczekiwaną możemy obliczyć wartość błędu. Robimy to przy pomocy funkcji błędu średniokwadratowego.

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

gdzie:

- $E(o)$ to błąd średniokwadratowy dla konkretnego neuronu wyjściowego
- n to liczba neuronów warstwy wyjściowej
- d to oczekiwane wyjście neuronu
- o to rzeczywiste wyjście neuronu

Dla sieci z rysunku 2 i wejścia (0,1) wartość jedyne neuronu wyjściowego jest równa 0.697. Natomiast jego wartość oczekiwana to 1. W związku z tym błąd średniokwadratowy jest równy:

$$E_o = 1/1 * (1 - 0.697)^2 \approx 0.092$$

Naszym celem będzie takie ustawienie wag, aby jak najbardziej zminimalizować błąd. Robimy to za pomocą wzoru:

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

gdzie:

w^+ jest nową wartością wagi.

w jest aktualną wartością wagi.

η jest stałą uczenia, czyli wartością, która określa szybkość uczenia sieci.

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

Szczególną uwagę należy poświęcić stałej uczenia η . 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 bądź równe 0.1. W naszym przypadku będzie to 0.1.

Kolejną rzeczą wartą uwagi jest pochodna błędu całkowitego do konkretnej wagi. Liczymy ją z wykorzystaniem reguły łańcuchowej, w przypadku w_{1o} :

$$\frac{\partial E_{total}}{\partial w_{1o}} = \frac{\partial E_{total}}{\partial out_o} * \frac{\partial out_o}{\partial net_o} * \frac{\partial net_o}{\partial w_{1o}}$$

Już teraz możemy zauważyć, że wzór na pochodną cząstkową dla w_{2o} będzie bardzo podobny:

$$\frac{\partial E_{total}}{\partial w_{2o}} = \frac{\partial E_{total}}{\partial out_o} * \frac{\partial out_o}{\partial net_o} * \frac{\partial net_o}{\partial w_{2o}}$$

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:

$$\delta_o = \frac{\partial E_{total}}{\partial net_o} = \frac{\partial E_{total}}{\partial out_o} * \frac{\partial out_o}{\partial net_o}$$

tym samym:

$$\begin{aligned} \frac{\partial E_{total}}{\partial w1o} &= \delta_o * \frac{\partial net_o}{\partial w1o} \\ \frac{\partial E_{total}}{\partial w2o} &= \delta_o * \frac{\partial net_o}{\partial w2o} \end{aligned}$$

Spróbujmy teraz uprościć wyrażenie $\frac{\partial net_o}{\partial w1o}$:

$$\frac{\partial net_o}{\partial w1o} = \frac{\partial (w1o * out_{h1} + w2o * out_{h2})}{\partial w1o} = 1 * out_{h1} * w1o^{(1-1)} + 0 + 0 = out_{h1}$$

Jak widzimy $\frac{\partial net_o}{\partial w1o}$ ostatecznie sprowadza się do wartości neuronu źródłowego dla połączenia. W analogiczny sposób upraszcza się równanie dla $\frac{\partial net_o}{\partial w2o}$, dzięki czemu otrzymujemy:

$$\begin{aligned} \frac{\partial E_{total}}{\partial w1o} &= \delta_o * out_{h1} \\ \frac{\partial E_{total}}{\partial w2o} &= \delta_o * out_{h2} \end{aligned}$$

Ostatnim krokiem potrzebnym do wyliczenia nowych wag jest obliczenie samego współczynnika delty dla o_1 (δ_o):

$$\delta_o = \frac{\partial E_{total}}{\partial net_o} = \frac{\partial E_{total}}{\partial out_o} * \frac{\partial out_o}{\partial net_o}$$

$$\frac{\partial E_{total}}{\partial out_o} = \frac{\partial (d_o - out_o)^2}{\partial out_o} = -(target_o - out_o)$$

$$\frac{\partial out_o}{\partial net_o} = \frac{\partial (\frac{1}{1+e^{-net_o}})}{\partial net_o} = (\frac{1}{1+e^{-net_o}})(1 - \frac{1}{1+e^{-net_o}}) \quad (1)$$

gdzie:

$$\frac{1}{1+e^{-net_o}} = out_o \quad (2)$$

a więc:

$$\frac{\partial out_o}{\partial net_o} = out_o(1 - out_o) \quad (3)$$

W związku z powyższymi równaniami wzór na współczynnik delty dla out_o wynosi:

$$\delta_o = \frac{\partial E_{total}}{\partial net_o} = \frac{\partial E_{total}}{\partial out_o} * \frac{\partial out_o}{\partial net_o} = -(target_o - out_o) * out_o(1 - out_o)$$

Co podstawieniu daje:

$$\delta_o \approx -0.064$$

W ten sposób otrzymaliśmy łatwy wzór na obliczenie pochodnych cząstkowej po $w1o$ i $w2o$

$$\frac{\partial E_{total}}{\partial w1o} = \delta_o * out_{h1} \approx -0.064 * 0.525 \approx -0.0336$$

$$\frac{\partial E_{total}}{\partial w2o} = \delta_o * out_{h2} \approx -0.064 * 0.690 \approx -0.04416$$

co umożliwia nam obliczenie nowych wag:

$$w1o^+ = w1o - \eta * \frac{\partial E_{total}}{\partial w1o} = 0.4 - 0.1 * -0.0336 = 0.40336$$

$$w2o^+ = w2o - \eta * \frac{\partial E_{total}}{\partial w2o} = 0.9 - 0.1 * -0.04416 = 0.904416$$

Spróbujmy teraz ustalić jakie wartości powinny przyjąć wagi połączeń pomiędzy warstwą wejściową a warstwą ukrytą. Stosujemy tutaj ten sam wzór co do połączeń pomiędzy warstwą ukrytą a warstwą wyjściową, tak więc wzór na wagę $w11$ będzie wyglądał tak:

$$w11^+ = w11 - \eta * \frac{\partial E_{total}}{\partial w11}$$

Tak jak poprzednio najbardziej problematyczne będzie wyznaczenie pochodnej błędu całkowitego do tej wagi:

$$\frac{\partial E_{total}}{\partial w11} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w11}$$

Od razu zauważmy, że pochodna błędu całkowitego do wagi drugiego połączenia (w21) do tego neuronu jest bardzo podobna:

$$\frac{\partial E_{total}}{\partial w_{21}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_{21}}$$

Podobnie jak w przypadku neuronów wyjściowych oznaczmy część wspólną tych równań jako $\delta h1$:

$$\delta h1 = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}}$$

Można jednak zauważyć, że pierwsza część równania to tak naprawdę δo , którą mamy już wyliczoną:

$$\delta h1 = \delta o * \frac{\partial net_{o1}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}}$$

Obliczmy teraz dwie pozostałe pochodne:

$$\frac{\partial net_{o1}}{\partial out_{h1}} = \frac{\partial (w_{1o} * out_{h1} + w_{2o} * out_{h2})}{\partial out_{h1}} = w_{1o}$$

$$\frac{\partial out_{h1}}{\partial net_{h1}} = \frac{\partial out_{h1}}{\partial net_{h1}} = out_{h1}(1 - out_{h1})$$

Tak więc ostatecznie wyznacznik delta dla neuronu h1 jest równy:

$$\delta h1 = \delta o * w_{1o} * out_{h1}(1 - out_{h1}) = -0.064 * 0.4 * 0.249375 = -0.006384$$

Wiedząc, że pochodna wartości neuronu przed funkcją aktywacji po wadze jest równa wartości neuronu z którego wychodzi połączenie możemy obliczyć również pochodne cząstkowe po wagach w11 i w21

$$\frac{\partial E_{total}}{\partial w_{11}} = \delta h1 * \frac{\partial net_{h1}}{\partial w_{11}} = \delta h1 * out_{i1} = -0.006384 * 0 = 0$$

$$\frac{\partial E_{total}}{\partial w_{21}} = \delta h1 * \frac{\partial net_{h1}}{\partial w_{21}} = \delta h1 * out_{i2} = -0.006384 * 1 = -0.006384$$

Dzięki którym jesteśmy w stanie obliczyć nowe wagi w11 i w21:

$$w_{11}^+ = w_{11} - \eta * \frac{\partial E_{total}}{\partial w_{11}} = 0.5 - 0.1 * 0 = 0.5$$

$$w_{21}^+ = w_{21} - \eta * \frac{\partial E_{total}}{\partial w_{21}} = 0.1 - 0.1 * -0.006384 = 0.1006384$$

W analogiczny sposób wyliczamy również wagi dla połączeń między neuronami warstwy wejściowej a neuronem h2:

$$w_{12}^+ = w_{12} - \eta * \frac{\partial E_{total}}{\partial w_{11}} = w_{12} - \eta * \delta h_2 * out_{i1} = 0.3 - 0.1 * -0.00547584 * 0 = 0.3$$

$$w_{22}^+ = w_{22} - \eta * \frac{\partial E_{total}}{\partial w_{21}} = w_{12} - \eta * \delta h_2 * out_{i2} = 0.8 - 0.1 * -0.00547584 * 1 = 0.800547584$$

Wielokrotne uruchomienie takiego algorytmu sprawi, że sieć będzie dawała oczekiwane rezultaty

5 Implementacja wielowarstwowej sieci neuronowej

Teraz przedstawię autorską implementację przedstawionego wcześniej algorytmu. Podstawą jest klasa odwzorowującej sieć neuronową, w moim 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<HiddenNeuron> HiddenLayer { get; }
5     IEnumerable<OutputNeuron> 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: Wyliczanie 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: Wyliczanie wyjścia połączenia

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