

Sprawozdanie  
Wielowarstwowe sieci neuronowe  
Amadeusz Filipek  
Laboratorium komputerowe WFIS AGH

1. Wstęp

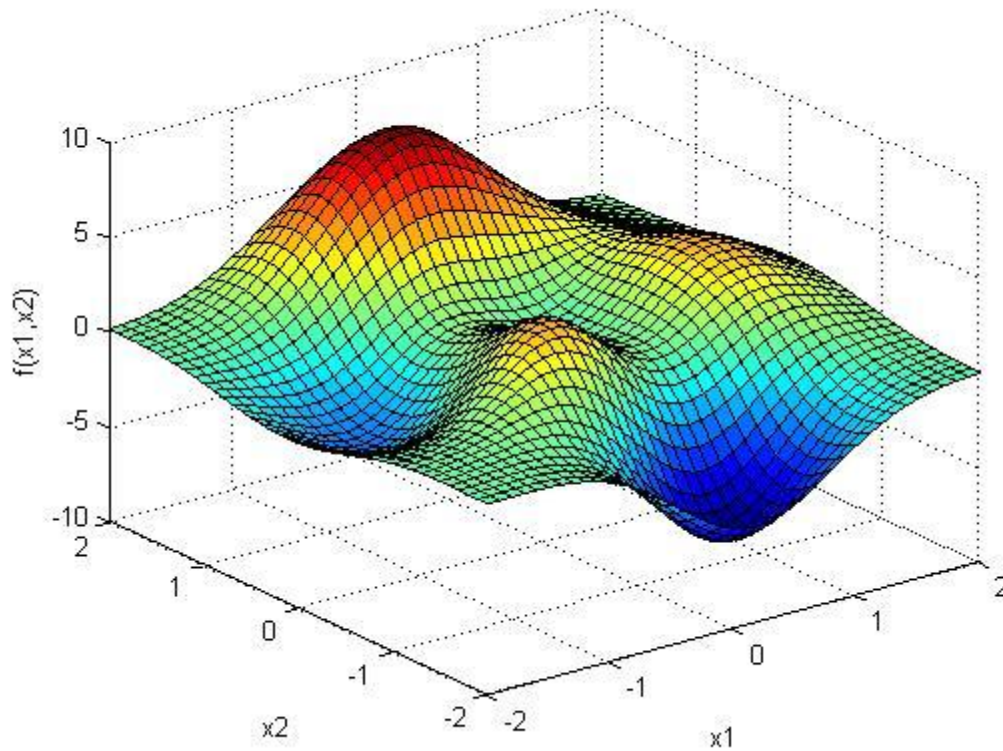
Celem ćwiczenia jest zastosowanie sieci neuronowych o różnej architekturze do rozwiązania problemu aproksymacji oraz klasyfikacji.

Sztuczna sieć neuronowa to rodzaj matematycznego modelu, który ma za zadanie przetwarzać informacje. Sieci neuronowe wchodzi w skład dziedziny nauki określonej jako nauczanie maszynowe (machine learning). Sieć neuronowa zbudowana jest z połączonych ze sobą neuronów ułożonych w kolejne warstwy. Taka sieć posiada warstwę neuronów wejściowych, określoną ilość warstw ukrytych oraz warstwę wyjściową. Sieć o najprostszej konfiguracji to wielowarstwowa sieć jednokierunkowa (MLP – Multi-Layer Perceptron). W tej konfiguracji każdy neuron poprzedniej warstwy połączony jest z wszystkimi neuronami warstwy kolejnej. Architektura sieci określona jest poprzez ilość warstw ukrytych, ilość neuronów w każdej warstwie ukrytej oraz funkcje aktywacji neuronów każdej warstwy. Niestety nie ma reguł, które określają jaka powinna być architektura sieci aby w sposób optymalny rozwiązywać dany problem. Kluczowym elementem determinującym sprawność sieci oprócz architektury jest proces nauczania. Typowym procesem nauczania w sieciach jednokierunkowych jest algorytm propagacji wstecznej błędu. Proces uczenia odbywa się na podstawie zbioru uczącego. Wagi każdego neuronu aktualizowane są na podstawie różnicy wyniku sieci i wartości przewidywanej w zbiorze uczącym. Efektywność otrzymanej sieci badana jest za pomocą osobnego zbioru testującego, proces ten określany jest mianem walidacji. Najprostsza walidacja polega na obliczeniu względnej różnicy pomiędzy wynikami sieci dla danych testujących a poprawnymi, znanymi wynikami. Wadą takiego podejścia jest potrzeba podzielenia posiadanego zbioru danych na zbiór uczący i walidujący, zatem algorytm uczący korzysta tylko z ułamka wiedzy. Bardziej zaawansowaną metodą jest walidacja krzyżowa, która polega na podzieleniu w sposób losowy zbioru danych na  $k$  równolicznych podzbiorów. Następnie sieć uczona jest na wszystkich zbiorach oprócz jednego, który wykorzystywany jest do wyliczenia błędu. Procedura ta powtarzana jest  $k$  razy, za każdym razem biorąc inny podzbiór do wyliczenia błędu. W ten sposób uzyskiwany jest zbiór wartości błędu na podstawie którego można wyznaczyć parametry statystyczne błędu działania sieci takie jak wartość średnia oraz odchylenie standardowe. Modyfikacją metody krzyżowej jest procedura walidacyjna leave-one-out, dla której  $k$  równe jest ilości danych w zbiorze uczącym.

## 2. Realizacja ćwiczenia – aproksymacja

W ćwiczeniu pierwszym sieć neuronową wykorzystano do aproksymacji funkcji dwuargumentowej o wzorze:

$$f(x_1, x_2) = 3(1 - x_1)^2 \exp(-(x_1^2) - (x_2 + 1)^2) - 10(x_1/5 - x_1^2 - x_2^2) \exp(-x_1^2 - x_2^2) - 1/3 \exp(-(x_1 + 1)^2 - x_2^2)$$



Wykres 1. Wykres funkcji aproksymowanej

Przygotowano 100 – elementowy zestaw uczący oraz 1000 – elementowy zestaw testujący. Oba zestawy zostały uzyskane losując w sposób jednorodny punkty  $(x_1, x_2)$  w zakresie  $[-3, 3]$ . W ćwiczeniu wykorzystano sieć neuronową z jedną warstwą ukrytą. Przetestowano architekturę sieci, próbując po 10, 20, 40 i 50 neuronów w warstwie ukrytej. Zbadano także jak zmienia się sprawność sieci w zależności od wykorzystanych funkcji aktywacji. W ćwiczeniu wykorzystano funkcje: skokową, liniową, hiperbolicznego tangensa sigmoidalnego oraz log-sigmoidalną. Maksymalna ilość epok uczenia ustalona została jako 200 epok. Sprawność sieci badano na podstawie błędu średniokwadratowego MSE:

$$MSE = \frac{1}{m} \sum_{i=1}^m (f_i - \hat{f}_i)^2$$

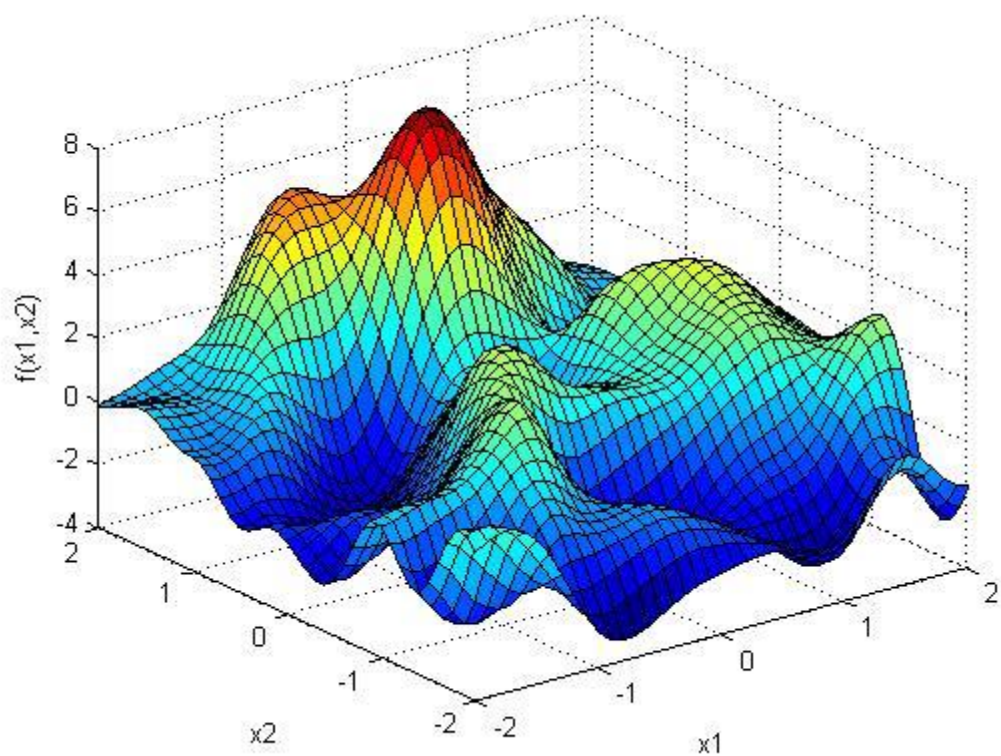
gdzie  $m$  – ilość elementów testowanego zbioru. Poniżej przedstawiono wyniki w zależności od wykorzystanej funkcji aktywacji oraz liczby neuronów ukrytych:

Tabela 1. Błąd średniokwadratowy sieci dla zbadanych architektur sieci

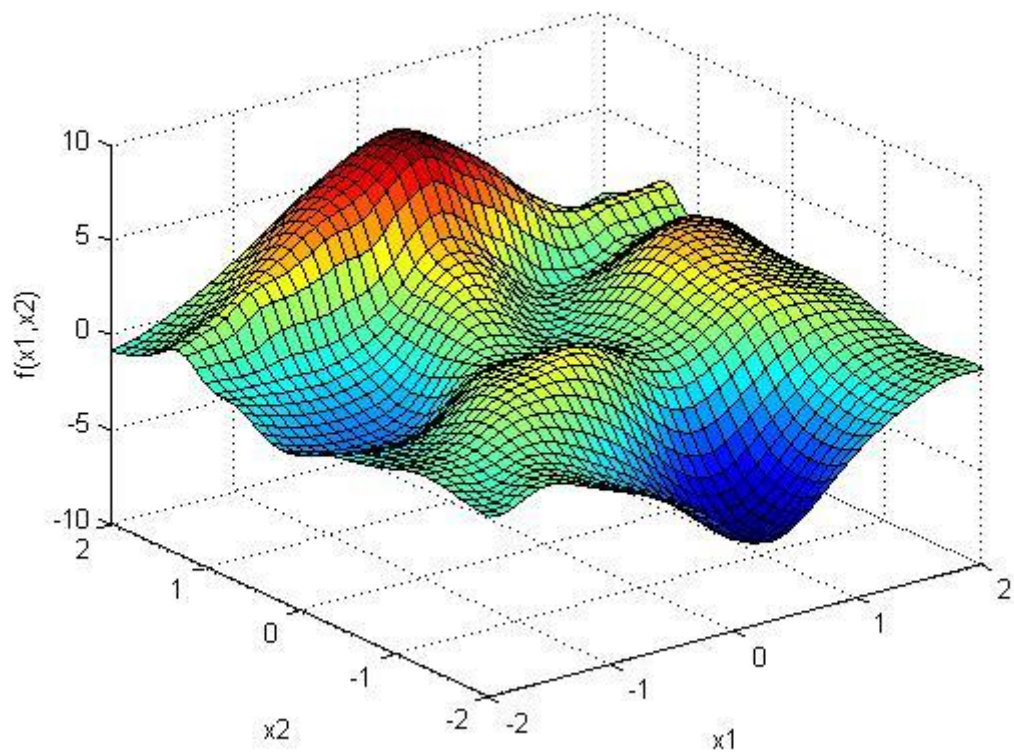
tansig	N neuronów	50	40	20	10
	MSE_ucz	2.1E-17	2.43E-08	0.77	0.86
	MSE_test	1.239	5.14	0.0014	0.1214
logsig	N neuronów	50	40	20	10
	MSE_ucz	2.15E-24	1.95E-23	0.012	0.2
	MSE_test	4.83	3.09	1.79	1.8
hardlim	N neuronów	50	40	20	10
	MSE_ucz	0.44	0.93	1.9	2.26
	MSE_test	2.02	3.84	2.32	2.7
purelin	N neuronów	50	40	20	10
	MSE_ucz	4.14	2.35	2.7	4.15
	MSE_test	3.12	3.05	3.06	2.99

Na podstawie powyższej tabeli można stwierdzić, że najlepsze rezultaty na zbiorach testujących uzyskane są przy użyciu funkcji hiperbolicznego tangensa sigmoidalnego, podczas gdy sieć uzyskuje najgorsze rezultaty przy użyciu funkcji liniowej. Zwiększenie ilości neuronów prowadzi do mniejszego błędu na zbiorze uczącym lecz wcale nie wykazuje tendencji do zmniejszenia błędu na zbiorze testującym. Najmniejsza wartość błędu na zbiorze uczącym została uzyskana dla sieci posiadającej 50 neuronów ukrytych o funkcji aktywacji log-sigmoidalnej, natomiast najmniejsza wartość błędu na zbiorze uczącym została uzyskana dla sieci posiadającej 20 neuronów ukrytych o funkcji aktywacji hiperbolicznego tangensa sigmoidalnego. Na poniższych wykresach przedstawiono wykresy funkcji aproksymowanej przez sieci o dwóch minimalnych wartościach błędów.

Na wykresie 2. widać, że funkcja jest mocno zafalowana, zwłaszcza na brzegach, a także posiada więcej ostrych wzniesień. Funkcja na wykresie 3. dobrze odwzorowuje funkcję przedstawioną na wykresie 1., trzy oryginalne wzniesienia są dobrze odwzorowane z nieznacznymi odchyleniami. Wciąż widoczne są zafalowania na brzegach.



Wykres 2. Wykres funkcji aproksymowanej przez nauczoną sieć o 50 neuronach ukrytych i funkcji aktywacji logsig



Wykres 3. Wykres funkcji aproksymowanej przez nauczoną sieć o 20 neuronach ukrytych i funkcji aktywacji tansig

### 3. Realizacja ćwiczenia – klasyfikacja

Drugie ćwiczenie polegało na wykorzystaniu jednokierunkowej sieci neuronowej w problemie klasyfikacji. Tym razem nauczanie odbywa się na danych parametrach fizykochemicznych napojów w celu klasyfikacji jakiej substancji konserwującej należy do danej próbki dodać. W rozważanym problemie korzysta się z trzech środków konserwujących zatem klasyfikacja obejmuje trzy klasy oznaczone liczbami 1, 2, 3. Na funkcję aktywacji neuronów ukrytych została wybrana funkcja log-sigmoidalna. Ze względu na ciągłą funkcję aktywacji, wyniki wyjściowe są liczbami rzeczywistymi, spośród nich wybierana jest największa a neuron wyjściowy odpowiadający tej wielkości decyduje o klasyfikacji.

Dobrano liczbę neuronów sieci ukrytej tak aby liczba błędnych klasyfikacji dla zbioru uczącego była minimalna. Zbiór uczący zawiera 130 elementów podczas gdy zbiór testujący 18 elementów. Minimalny błąd uzyskano dla sieci o 20 neuronach i wynosił on 0% błędnych klasyfikacji dla zestawu uczącego. Następnie przetestowano pracę tak nauczanej sieci na zbiorze testującym i uzyskano wynik 33% błędnych klasyfikacji.

Kolejnym etapem ćwiczenia była zmiana funkcji aktywacji neuronów ukrytych na funkcję skokową *hardlim* przy zachowaniu architektury sieci. W tym wypadku wartości wyjściowe sieci są liczbami z zakresu [0, 1] zatem interpretacja wyników jest łatwiejsza. Należy jedynie uwzględnić sytuację w której na wyjściu uzyskiwane są dwie lub więcej jedynek, wtedy klasyfikacja jest niejednoznaczna. Uzyskano 6.9% błędnych klasyfikacji dla zbioru uczącego oraz 27.8% błędnych klasyfikacji dla zbioru testującego.

Uzyskane wyniki pokazują, że sieć pracująca na funkcji aktywacji log-sigmoidalnej bezbłędnie klasyfikuje zbiór uczący, natomiast dla zbioru testującego sieć uzyskuje dość duży błąd co może być przejawem przeuczenia sieci. Przejście do funkcji skokowej powoduje zwiększenie błędu dla zbioru uczącego i nieznaczną poprawę dla zbioru testującego. Zwiększony błąd dla zbioru uczącego może wynikać z problemu interpretacji dwóch lub więcej jedynek na wyjściu podczas gdy dla funkcji ciągłej taka sytuacja nie zachodzi (zawsze jeden neuron ma wartość największą). Duży błąd dla zbioru testującego w obu przypadkach może być związany z faktem że zbiór testujący jest mały i zawiera 18 elementów.

### 4. Podsumowanie

W ramach ćwiczenia wykorzystano sieć neuronową w problemach aproksymacji oraz klasyfikacji i sprawdzono jak zmienia się sprawność sieci z zależności od wykorzystanej architektury oraz funkcji aktywacji. Dwuargumentowa funkcja została dobrze odtworzona przez sieć uzyskując wartość 0.0014 błędu średniokwadratowego przy konfiguracji 20 neuronów w sieci ukrytej o funkcji aktywacji w postaci hiperbolicznego tangensa sigmoidalnego. Większa ilość neuronów wcale nie prowadziła do lepszej aproksymacji szukanej funkcji ze względu na efekt przeuczenia sieci. Co więcej, nieliniowe funkcje aktywacji znacznie lepiej radzą sobie z odwzorowaniem nieliniowej funkcji. Należy zaznaczyć, że dane uczące zostały wygenerowane w sposób jednorodny w całym obszarze funkcji, zatem gęstość informacji o aproksymowanej funkcji jest w przybliżeniu jednorodna co przekłada się na dobry wynik sieci. Problem klasyfikacji trójklasowej został rozwiązany z wynikiem 27.8% błędnych klasyfikacji zbioru testującego dla dyskretnej funkcji aktywacji oraz z 33% błędem dla ciągłej funkcji aktywacji. Słaba sprawność sieci na zbiorze testującym może być tłumaczona faktem, że zbiór testujący zawierał tylko 18 elementów w stosunku do 130 elementów zbioru testującego.