

Aproksymacja funkcji charakterystycznej trójkąta przy użyciu sieci neuronowych.

Bartłomiej Chęciński, Kuba Buraczewski

29 marca 2025

1 Wstęp

Celem zadania jest aproksymacja funkcji charakterystycznej trójkąta o wierzchołkach $(0, 0)$, $(1, 0)$, $(0, 1)$ przy użyciu sieci neuronowych o różnych architekturach. Dane treningowe generowane są z kwadratu $[0, 1] \times [0, 1]$. Każdy punkt otrzymuje etykietę 1, jeśli leży w obszarze trójkąta, w przeciwnym przypadku 0.

2 Proces treningu

W celu nauczania modeli zastosowano:

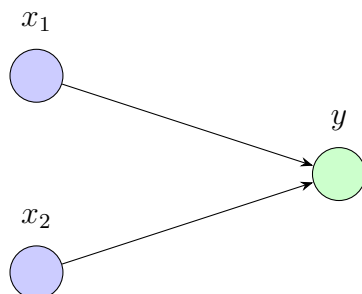
- **Funkcję straty MSE** (błąd średniokwadratowy).
- **Optymalizator Adam**, który iteracyjnie aktualizuje wagi.
- **Wczesne zatrzymanie** – jeśli błąd spadnie poniżej ustalonego progu, trening kończy się przed osiągnięciem maksymalnej liczby epok.

Testy przeprowadzono dla trzech popularnych funkcji aktywacji: **ReLU**, **Tanh** i **Sigmoid**.

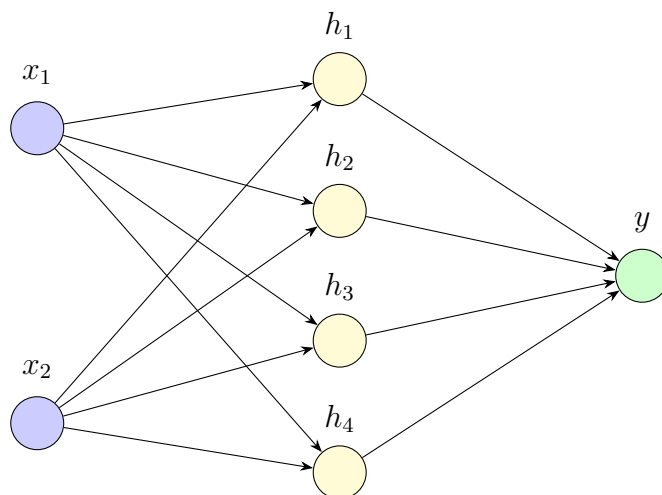
3 Architektury sieci

Przygotowano trzy rodzaje sieci neuronowych, różniące się głębokością. Poniżej przedstawiono ich poglądowe rysunki. Na rysunkach pominięto biasy.

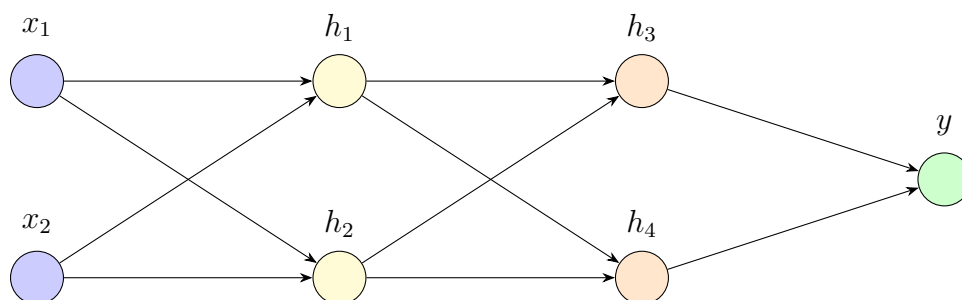
Sieć jednowarstwowa (3 parametry)



Sieć dwuwarstwowa (17 parametrów)



Sieć trzwarstwowa (15 parametrów)



4 Wyniki i wnioski

Dla każdego modelu przeprowadzono eksperymenty z różnymi funkcjami aktywacji (ReLU, Tanh, Sigmoid). Wyniki pokazały, że:

- Czym głębsza sieć tym lepiej radziła sobie z aproksymacją.
- Czasami pojawiały się problemy z funkcją ReLU, jest to powszechny problem tzw. dying ReLU. ReLU zwraca zero dla wszystkich wartości wejściowych mniejszych od zera i w tym zakresie jej pochodna też wynosi zero. W efekcie, jeśli podczas inicjalizacji (albo bardzo szybko w trakcie uczenia) sieć ustawi się tak, że $w \cdot x + b < 0$ dla wszystkich punktów w warstwie, to wyjście będzie wszędzie równe 0, a gradienty – wszędzie zerowe. Model „zastyga” wtedy w rozwiązaniu zwracającym wyłącznie zero i nie ma już mechanizmu, który wyciągnąłby go z tego stanu.
- Modele były zdolne do generalizacji na nowe dane.

5 Podsumowanie

W ramach zadania zbadano wpływ różnych głębokości sieci neuronowych oraz wyboru funkcji aktywacji na jakość aproksymacji funkcji charakterystycznej trójkąta. Uzyskane wyniki pokazują, że większa liczba warstw może poprawiać jakość aproksymacji, choć kosztem większej złożoności obliczeniowej.