Podstawy Sztucznej Inteligencji

Scenariusz 2

Budowa i działanie neuronowej sieci jednowarstwowej**.**

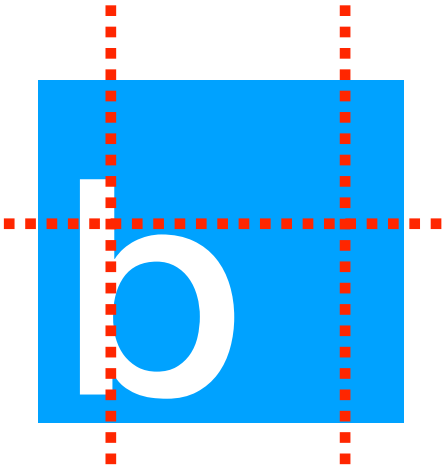
Cel ćwiczenia

Poznanie budowy i działania jednowarstwowych sieci neuronowych oraz uczenie rozpoznawania wielkości liter.

Założenia

Aby uczenie rozpoznawania wielkości liter w sieci jednowarstwowej było możliwe, musimy przydzielić neuronom pewne podzadania, których wyniki będą przetwarzane przez osobny neuron. Podzieliłem problem rozpoznawania wielkości litery na podzadania:

1. Czy litera przekracza lewą linię?
2. Czy litera przekracza prawą linię?
3. Czy litera przekracza górną linię?



Większość dużych litery takich jak „A” lub „O” przekracza wszystkie trzy linie, lecz np. „I” przekracza tylko górną linię, a „L” lewą oraz górną (zakładam, że musi przekraczać większością swojego obszaru). Widać, że tutaj pojawia się pewne skomplikowanie, które będzie idealnym zadaniem dla sztucznej inteligencji.

Zatem w warstwie będą znajdować się trzy neurony sigmoidalne lub Adaline (o tym więcej w dalszej części), gdzie każdy z nich będzie uczył się rozpoznawać jedno z trzech wyżej wymienionych podzadań. W czasie uczenia tej warstwy, osobny perceptron będzie uczył się rozpoznawać czy litera jest duża czy mała na podstawie oczekiwanych odpowiedzi dla warstwy. Oczywiście perceptron nie będzie uczył się na podstawie odpowiedzi neuronów warstwy, ponieważ potrzebowalibyśmy zaimplementować algorytm wstecznej propagacji błędu.

Neuron sigmoidalny

To taki neuron, którego funkcja w przeciwieństwie do modelu McCullocha-Pitsa jest ciągła i przyjmuje postać funkcji sigmoidalnej (w tym przypadku unipolarnej):

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

Argumentem dla tej funkcji jest oczywiście sygnał sumacyjny *u.* Współczynnik beta jest dobierany przez programistę i wpływa on na kształt funkcji. Zazwyczaj jednak używa się beta równego 1 i również tak jest w mojej sieci. Ważną cechą funkcji sigmoidalnej jest jej różniczkowalność, dzięki której **możemy zastosować metodę gradientową**. Najprościej jest

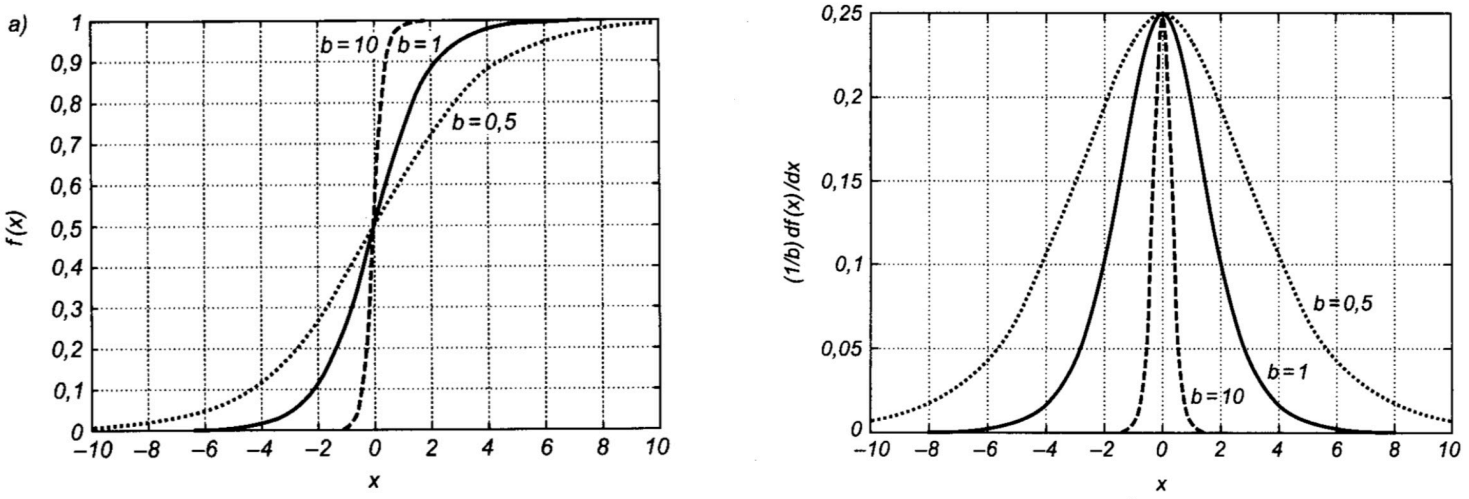
przyjąć metodę największego spadku, zgodnie z którą aktualizacja wektora wag odbywa się w kierunku ujemnego gradientu funkcji celu. Uaktualnianie w sposób dyskretny:

$$w_{ij}(k+1) = w_{ij}(k) - \eta\,\delta_i\,x_j$$

gdzie: $$\delta_i = e_i \frac{\mathrm{d}f(u_i)}{\mathrm{d}u_i}$$

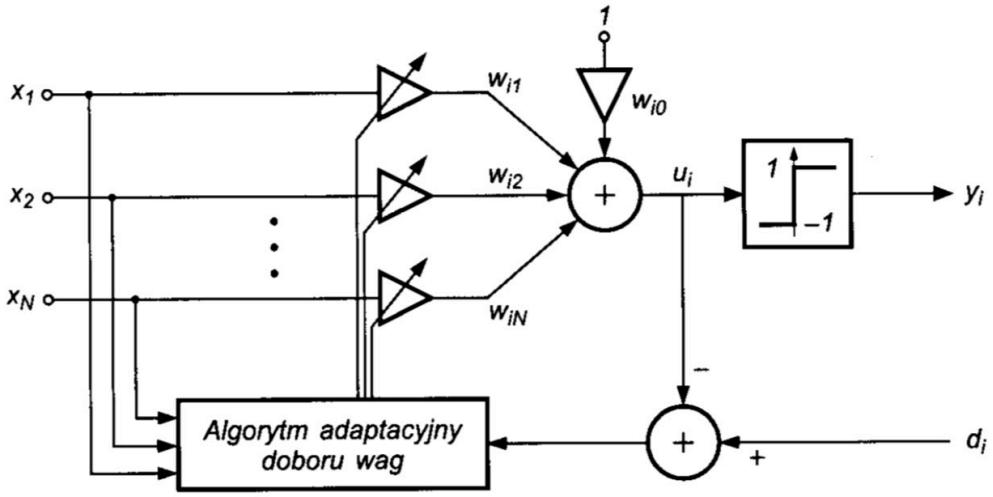
$$e_i = (y_i - d_i)$$

Wykres funkcji sigmoidalnej oraz jej pochodnej:



Neuron Adaline (Adaptive Linear Neuron)

Opracowany przez B.Widrowa. Funkcja aktywacji jest typu signum a jego schemat adaptacyjnego sposobu doboru wag przedstawia poniższy rysunek:



Rys. 2.6. Schemat neuronu typu adaline

Główną różnicą Adaline a np. Perceptronem (który też posiada model nieliniowy) jest to, że **w definicji funkcji celu używa jedynie części liniowej** (sumę wagową sygnałów wejściowych). Dzięki temu właśnie jest możliwe zastosowanie algorytmu gradientowego uczenia. W minimalizacji funkcji celu używa się metodę największego spadku, podobnie jak w przypadku neuronu sigmoidalnego.

$$w_{ij}(k+1) = w_{ij}(k) + \eta\, e_i\, x_j$$

gdzie: $$e_i = \left(d_i - \sum_{j=0}^{N} w_{ij} x_j\right)$$

Wyniki

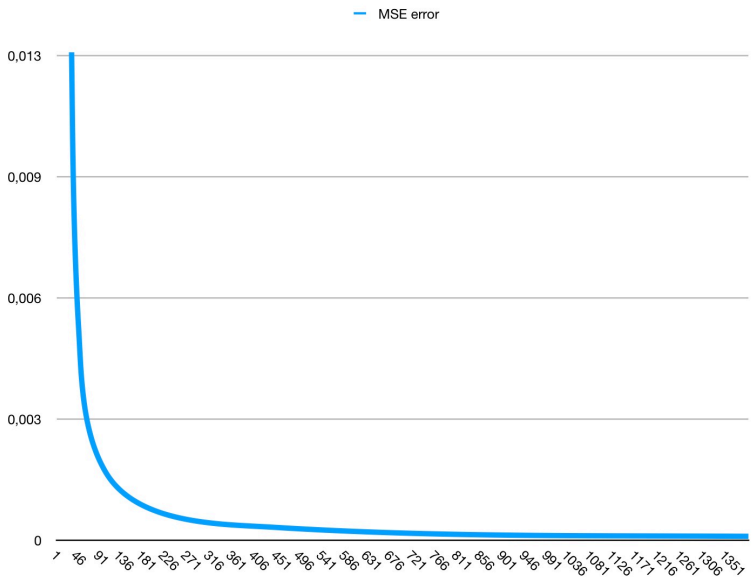
Rezultaty były zaskakująco dobre jak na jedną tylko warstwę i czasem subiektywny podział rozpoznawania wielkości na podzadania. W pętli uczenia sieci dodałem warunek kończący ją w momencie uzyskania błędu średniokwadratowego mniejszego niż 0.001. Przeprowadziłem testy warstwy neuronów sigmoidalnych dla współczynnika uczenia równego **0.1, 0.5, 0.01 oraz 0.05.** Dokładne wyniki zapisane są w katalogu /Charts/. Zgodnie z moimi przewidywaniami okazało się, że współczynnik uczenia wpływa mocno na szybkość uczenia się warstwy.

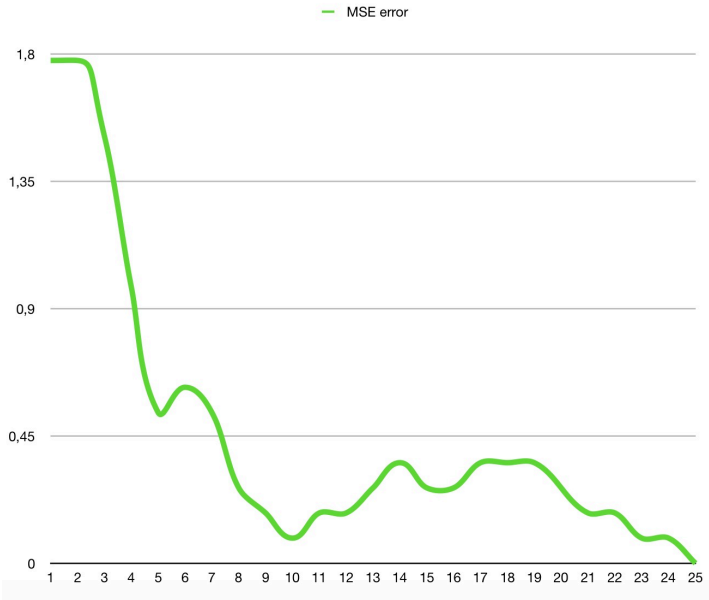
- Dla lRate = **0.01** - sieć nauczyła się po upływie **64414** epok
- Dla lRate = **0.05** - sieć nauczyła się po upływie **13420** epok
- Dla lRate = **0.1** - sieć nauczyła się po upływie **6816** epok
- Dla lRate = **0.5** - sieć nauczyła się po upływie **1315** epok

Dla Adaline nie udało mi się przeprowadzić wszystkich testów, gdyż dla lRate równego 0.1 i 0.5 uczenie stawało w miejscu na MSE równym około 0.24. Po długiej analizie doszedłem do wniosku, że ma to związek z danymi wejściowymi, które dla takiego dużego lRate ujawniają podatność Adaline na zatrzymywanie się w miejscu podczas uczenia, spowodowaną taką a nie inną budową funkcji celu (sposobem aktualizacji wag z wykorzystaniem tylko sumy wagowej). Udało się natomiast uzyskać wyniki dla:

- dla lRate - **0.01** - sieć nauczyła się po upływie **39** epok
- dla lRate - **0.05** - sieć nauczyła się po upływie **25** epok

Najszybsze uczenie sieci **Sigmoidalnej** vs. **Adaline:**





Spostrzeżenia

Po serii testów okazuje się, że Adaline źle radzi sobie z rozpoznawaniem nieznanych jej liter. Natomiast warstwa sigmoidalna z nieznanymi literami nie ma większego problemu. Niestety jak widać wyżej warstwa sigmoidalna potrzebuje więcej czasu aby się nauczyć na podstawie danych treningowych.

Wnioski

W tak prostym przykładzie jak rozpoznawanie wielkości oraz w prostej sieci jednowarstwowej składającej się z tylko trzech neuronów liter współczynnik uczenia bardzo widocznie zmienia szybkość uczenia się warstwy. Zatem im większy współczynnik uczenia, tym mniejszy czas uczenia. W tym prostym przykładzie w warstwie sigmoidalnej duży współczynnik uczenia zazwyczaj nie tworzy nawet wahań w wartości MSE. W Adaline pojawiają się małe wahania, lecz nadal nie ma to znaczenia przy tak małej ilości epok. Przy dużym współczynniku raz na jakiś czas uczenie staje w miejscu przez to, że aktualizacja wagi może „przekręcić" jej wartość w drugą stronę.

Ponadto okazuje się, że Adaline jest znacznie gorsze w przypadku, jeżeli chcemy testować litery, których sieć nie poznała podczas uczenia. Przez to, że Adaline tak szybko się uczy to dla dużych współczynników uczenia może nie potrafić się nauczyć.