

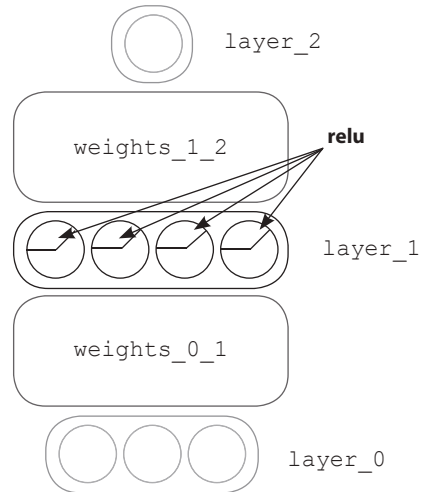
Czym jest funkcja aktywacji?

Jest to funkcja stosowana do neuronów pewnej warstwy podczas prognozowania.

Funkcja aktywacji jest funkcją stosowaną do neuronów warstwy podczas prognozowania. Powinno to wyglądać znajomo, gdyż używaliśmy już funkcji aktywacji o nazwie `relu` (pokazanej obok w trójwarstwowej sieci neuronowej). Funkcja `relu` powodowała zmianę wszystkich wartości ujemnych na 0.

Upraszczając, funkcja aktywacji to dowolna funkcja, która może przyjąć jedną wartość i zwrócić inną. Jednak liczba najróżniejszych istniejących funkcji jest nieskończona i zdecydowanie nie wszystkie nadają się do wykorzystania jako funkcje aktywacji.

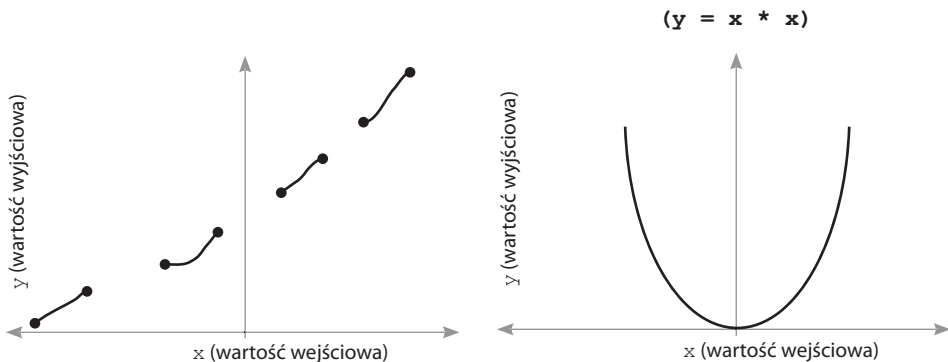
Istnieje kilka ograniczeń zapewniających, że jakaś funkcja może być funkcją aktywacji. Jak się przekonamy, używanie funkcji spoza tych ograniczeń jest zazwyczaj kiepskim pomysłem.



Ograniczenie 1: Funkcja musi być ciągła i mieć nieograniczoną dziedzinę.

Pierwsze ograniczenie budujące właściwą funkcję aktywacji jest takie, że musi ona zwracać wartość wyjściową dla *dowolnej* wartości wejściowej. Mówiąc inaczej, nie możemy być w stanie znaleźć takiej liczby, dla której z jakiegoś powodu nie istniałaby wartość tej funkcji.

To mała przesada, ale czy widać, że funkcja po lewej (cztery oddzielne linie) nie ma wartości y dla każdej wartości x ? Jest zdefiniowana tylko dla czterech przedziałów.



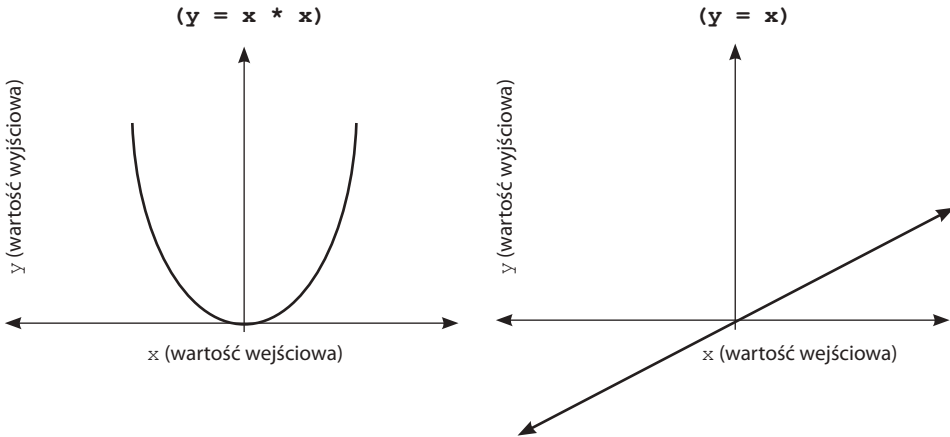
Byłaby to koszmarna funkcja aktywacji. Natomiast funkcja po prawej jest ciągła i ma nieograniczoną dziedzinę. Nie istnieje taka wartość wejściowa (x), dla której nie moglibyśmy obliczyć wyjścia (y).

Ograniczenie 2: Dobre funkcje aktywacji są monotoniczne, nigdy nie zmieniają kierunku.

Drugie ograniczenie mówi, że funkcja jest monotoniczna. Nigdy nie może zmienić kierunku. Innymi słowy, musi albo zawsze być nierosnąca, albo zawsze niemalejąca¹.

Jako przykład rozważmy poniższe dwie funkcje. Ich kształty odpowiadają na pytanie „Jeśli x to wejście, jaką wartość y opisuje ta funkcja?”. Funkcja po lewej ($y = x * x$) nie jest idealną funkcją aktywacji, gdyż nie jest stale nierosnąca ani stale niemalejąca.

Jak możemy to stwierdzić? Zauważmy, że mamy tu wiele przypadków, w których dwóm (różnym) wartościom x odpowiada pojedyncza wartość y (jest to prawdą dla każdej liczby z wyjątkiem 0). Natomiast funkcja po prawej jest zawsze rosnąca! Nie istnieje taki punkt, dla którego dwie wartości x dawałyby tę samą wartość y :



To konkretne ograniczenie nie jest wymagane z technicznego punktu widzenia. W przeciwieństwie do funkcji, dla których brakuje pewnych wartości (o nieciągłej dziedzinie), można optymalizować funkcje, które nie są monotoniczne. Trzeba jednak rozważyć implikacje sytuacji, gdy wiele wartości wejściowych przekształcanych jest na tę samą wartość wyjściową.

Podczas uczenia się sieci neuronowej poszukujemy właściwych konfiguracji wag, zapewniających określone wyjście. Problem ten może stać się znacznie trudniejszy, jeśli będzie istnieć wiele poprawnych odpowiedzi. Jeśli istnieje wiele sposobów

¹ W niektórych architekturach sieci neuronowych (np. RBF) możemy jednak spotkać niemonotoniczne funkcje aktywacji, lecz takie sieci wymagają specyficznego procesu uczenia (przyp. red.).