

Która z funkcji błędu (straty) stosowana jest najczęściej do problemów klasyfikacji?

Która z funkcji błędu (straty) stosowana jest najczęściej do problemów klasyfikacji?

- ☐ a. Średni błąd bezwzględny (mean absolute error, MAE)
- ☒ b. Entropia krzyżowa (cross-entropy) ✓
- ☐ c. Błąd średniokwadratowy (mean-squared error, MSE)

Na wejścia neuronu podany jest wektor 3 wartości $x=(1,0,5)$. Wagi przypisane do poszczególnych wejść są następujące $w=(0.5, 0.1, 0.7)$. Natomiast bias wynosi $b=-0.3$. Zastosowano funkcję aktywacji ReLU. Jaka będzie wartość na wyjściu neuronu?

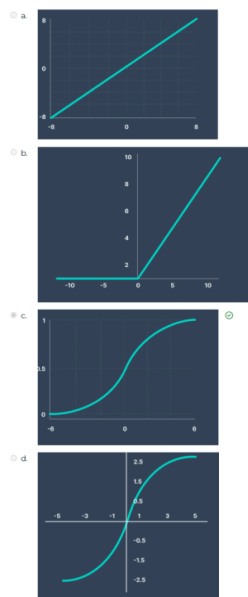
Na wejścia neuronu podany jest wektor 3 wartości $x=(1,0,5)$. Wagi przypisane do poszczególnych wejść są następujące $w=(0.5, 0.1, 0.7)$. Natomiast bias wynosi $b=-0.3$. Zastosowano funkcję aktywacji ReLU. Jaka będzie wartość na wyjściu neuronu?

- ☐ a. 4
- ☒ b. 3.7 ✓
- ☐ c. -0.3
- ☐ d. 0

Na wejścia neuronu podany jest wektor 3 wartości $x=(1,0,3)$. Wagi przypisane do poszczególnych wejść są następujące $w=(0.2, 0.1, -0.7)$. Natomiast bias wynosi $b=1.3$. Zastosowano funkcję aktywacji ReLU. Jaka będzie wartość na wyjściu neuronu?

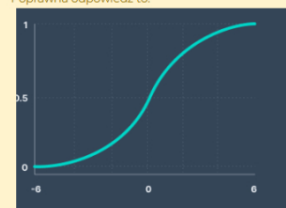
- ☐ a. 1
- ☒ b. 0 ✓
- ☐ c. 0.6
- ☐ d. -0.6

Który z poniższych wykresów przedstawia funkcję aktywacji sigmoid

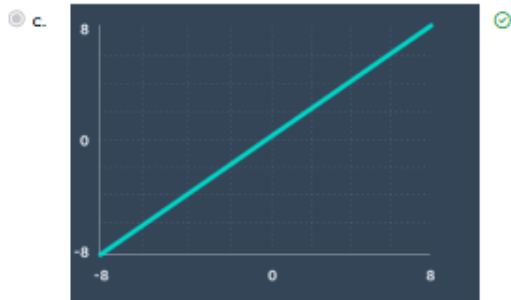


Twoja odpowiedź jest poprawna.

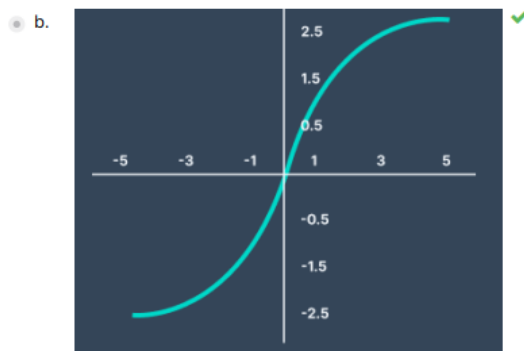
Poprawna odpowiedź to:



Który z poniższych wykresów przedstawia liniową funkcję aktywacji

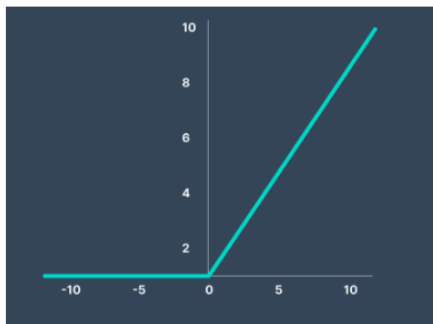


Który z poniższych wykresów przedstawia funkcję aktywacji TanH



Który z poniższych wykresów przedstawia funkcję aktywacji ReLU

☒ d.



Współczynnik (prędkość) uczenia:

Współczynnik (prędkość) uczenia:

- ☒ a. definiuje jak mocno zmieniają się wagi w kierunku zdefiniowanym przez algorytm wstecznej propagacji ✓
- ☐ b. definiuje kierunek aktualizacji wag
- ☐ c. nie ma wpływu na szybkość zbieżności uczenia
- ☐ d. zawsze jest większy lub równy 1.

W algorytmie wstecznej propagacji błędu:

W algorytmie wstecznej propagacji błędu:

- ☐ a. Zawsze wykorzystuje się błąd średniokwadratowy.
- ☒ b. Wagi aktualizuje się korzystając ze wzoru $w'_i = w_i - \mu \frac{\delta L}{\delta w_i}$ ✓
- ☐ c. Nie można zastosować wstecznej propagacji błędu do sieci konwolucyjnych.
- ☐ d. Aktualizacja wag następuje zawsze po każdym przykładzie uczącym.

W algorytmie wstecznej propagacji błędu:

- ☐ a. Żadna z pozostałych odpowiedzi nie jest poprawna.
- ☐ b. Wagi jednostek aktualizuje się zaczynając od pierwszej warstwy.
- ☐ c. Zmiana (aktualizacja) wag wszystkich jednostek jest jednakowa.
- ☒ d. Wagi jednostek aktualizuje się zaczynając od ostatniej warstwy. ✓

Stochastyczny spadek gradientu:

Stochastyczny spadek gradientu:

- ☐ a. wyznacza aktualizacje wag w oparciu o wszystkie przykłady uczące
- ☒ b. pozwala "uciec" z lokalnego optimum ✓
- ☐ c. w większości przypadków daje wyniki lepsze (bliższe optimum) niż klasyczny algorytm spadku gradientu.
- ☐ d. jest wolniejszy od klasycznego algorytmu spadku gradientu

Pytanie 3

Poprawnie

Punkty: 1,00 z 1,00

W algorytmie wstecznej propagacji błęd:

- ☐ a. Zmiana (aktualizacja) wag wszystkich jednostek jest jednakowa.
- ☐ b. Wagi jednostek aktualizuje się zaczynając od pierwszej warstwy.
- ☒ c. Wagi jednostek aktualizuje się zaczynając od ostatniej warstwy. ✓
- ☐ d. Żadna z pozostałych odpowiedzi nie jest poprawna.

Twoja odpowiedź jest poprawna.

Poprawna odpowiedź to:

Wagi jednostek aktualizuje się zaczynając od ostatniej warstwy.

Pytanie 16

Poprawnie

Punkty: 1,00 z 1,00

W algorytmie wstecznej propagacji błęd we wzorze na aktualizację wag: $w'_i = w_i - \mu \frac{\delta L}{\delta w_i}$

- ☒ a. μ oznacza prędkość uczenia. ✓
- ☐ b. μ oznacza wartość funkcji aktywacji neuronu.
- ☐ c. μ oznacza wartość funkcji straty (błęd).
- ☐ d. μ oznacza wartość wejściową neuronu dla aktualnego przykładu uczącego.

Twoja odpowiedź jest poprawna.

Poprawna odpowiedź to:

μ oznacza prędkość uczenia.

Dana jest sieć neuronowa składająca się z 3 warstw w pełni połączonych zawierających odpowiednio 64, 32 i 10 jednostek (neuronów). Pojedynczy przykład uczący składa się z 10 wartości. Ile zmiennych uczonych (tylko wag, bez biasów) znajduje się w tej sieci?

Dana jest sieć neuronowa składająca się z 3 warstw w pełni połączonych zawierających odpowiednio 64, 32 i 10 jednostek (neuronów). Pojedynczy przykład uczący składa się z 10 wartości. Ile zmiennych uczonych (**tylko wag, bez biasów**) znajduje się w tej sieci?

Odpowiedź: 1060



Poprawna odpowiedź to: 3008

Warstwa konwolucyjna:

Warstwa konwolucyjna:

- ☒ a. realizuje operacje splotu na wycinku danych wejściowych ✓
- ☐ b. jest warstwą nieuczalną
- ☐ c. wymaga znacznie większej liczby zmiennych uczonych niż warstwa w pełni połączona (gęsta).
- ☐ d. wymaga "spłaszczenia" danych wejściowych do jednego wymiaru

Padding w warstwach konwolucyjnych:

Padding w warstwach konwolucyjnych:

- ☐ a. nie wpływa na rozmiar danych wyjściowych neuronu
- ☒ b. definiuje w jaki sposób zachowuje się brzeg danych (np piksele brzegowe na obrazie) ✓
- ☐ c. definiuje o ile pól (np. pikseli) przesuwa się "okno" filtra
- ☐ d. w warstwach konwolucyjnych nie ma takiego parametru

Jaki jest cel stosowania regularyzacji w sieciach neuronowych?

Jaki jest cel stosowania regularyzacji w sieciach neuronowych?

Tekst odpowiedzi Pytanie 10

Regularyzację w sieciach neuronowych wykorzystuje się w celu uniknięcia przeuczenia sieci.

Jaki jest cel stosowania regularyzacji w sieciach neuronowych?

Tekst odpowiedzi Pytanie 11

Celem stosowania regularyzacji w sieciach neuronowych jest ograniczenie przeuczenia.

Technika batch normalization

Technika batch normalization

- ☐ a. spowalnia proces uczenia sieci
- ☐ b. nie jest mechanizmem regularyzacji.
- ☐ c. działa w ten sam sposób podczas uczenia sieci jak i w fazie wykorzystanie gotowego modelu.
- ☒ d. normalizuje wyjście warstwy tak, żeby średnia była równa ok. 0 a odchylenie standardowe ok 1. ✓

Warstwa dropout

Warstwa dropout

- ☐ a. stosowana jest zazwyczaj w małych sieciach
- ☐ b. nie jest mechanizmem walki z przeuczeniem
- ☒ c. powoduje tymczasowe "wyłączenie" (ignorowanie) niektórych neuronów ✓
- ☐ d. zawsze poprawia wyniki sieci neuronowej

Jaki typ sieci neuronowych sprawdzi się najlepiej do analizy cen papierów wartościowych na giełdzie w celu predykcji decyzji co do zakupu lub zbycia danych akcji?

Jaki typ sieci neuronowych sprawdzi się najlepiej do analizy cen papierów wartościowych na giełdzie w celu predykcji decyzji co do zakupu lub zbycia danych akcji?

- ☒ a. Sieci rekurencyjne ✓
- ☐ b. Sieci konwolucyjne
- ☐ c. Sieci warstwowe

Rekurencyjne sieci neuronowe wykorzystywane są najczęściej do:

Rekurencyjne sieci neuronowe wykorzystywane są najczęściej do:

- ☒ a. analizy przebiegów czasowych i danych tekstowych ✓
- ☐ b. żadna z pozostałych odpowiedzi nie jest prawdziwa
- ☐ c. analizy obrazów
- ☐ d. problemów prostej klasyfikacji

Rekurencyjne sieci neuronowe:

Rekurencyjne sieci neuronowe:

- ☐ a. sprawdzają się do przetwarzania obrazów
- ☐ b. nie wykorzystują funkcji aktywacji
- ☐ c. mają zawsze dwie warstwy
- ☒ d. przyjmują na wejściu wynik (wyjście) poprzedniej iteracji ✓

Zaznacz poprawną definicję neuronu i sieci neuronowej.

Zaznacz poprawną definicję neuronu i sieci neuronowej.

- ☒ a. Neuron to jednostka przetwarzająca dane w sieci neuronowej, która wykonuje operacje matematyczne, a sieć neuronowa to zbiór połączonych neuronów współpracujących w celu rozwiązywania problemów. ✓
- ☐ b. Neuron to algorytm klasyfikujący dane, a sieć neuronowa to model liniowy służący do predykcji.
- ☐ c. Neuron to fizyczny procesor komputerowy, a sieć neuronowa to układ sprzętowy służący do przetwarzania danych.
- ☐ d. Neuron to jednostka pamięci przechowująca dane, a sieć neuronowa to baza danych składająca się z neuronów.

Warstwy rekurencyjne:

Pytanie 7

Poprawnie

Punkty: 2,50 z 2,50

Warstwy rekurencyjne:

- ☐ a. Służą do przetwarzania obrazów
- ☐ b. Wykorzystywane są w sieci Resnet
- ☐ c. Nie mają uczalnych parametrów
- ☒ d. Służą do przetwarzania sekwencji ✓

Twoja odpowiedź jest poprawna.

Poprawna odpowiedź to: Służą do przetwarzania sekwencji

Zaznacz poprawną definicję enkodera i dekodera (autoenkoder)

Zaznacz poprawną definicję enkodera i dekodera (autoenkoder)

- ☐ a. Enkoder to model nadzorowany do klasyfikacji danych, a dekodery to model nienadzorowany do ich grupowania.
- ☐ b. Enkoder to algorytm przetwarzający dane na format binarny, a dekodery to proces odwrotny, który przywraca dane do ich pierwotnego formatu.
- ☒ c. Enkoder to część autoenkodera, która kompresuje dane wejściowe do mniejszej reprezentacji, a dekodery to część, która rekonstruuje dane wyjściowe z tej reprezentacji. ✓
- ☐ d. Enkoder i dekodery to elementy używane wyłącznie w przetwarzaniu języka naturalnego do tłumaczenia tekstu między językami.

Które ze stwierdzeń dotyczących nieliniowych funkcji aktywacji jest niepoprawne?

Które ze stwierdzeń dotyczących nieliniowych funkcji aktywacji jest **niepoprawne**?

- ☐ a. Zastosowanie nieliniowych funkcji aktywacji pozwala na dynamiczną selekcję danych poprzez "wylączenie" (wyzerowanie wyjść) niektórych neuronów.
- ☐ b. Zastosowanie funkcji aktywacji pozwala ograniczyć zakres wartości wyjściowych neuronu.
- ☒ c. Zastosowanie funkcji aktywacji ReLU ogranicza zakres wartości wyjściowych do przedziału [0-1] ✓
- ☐ d. Funkcje aktywacji stosuje się, aby sieć potrafiła modelować złożone (nieliniowe) zależności

Autoenkodery rozwiązujące problem autoasocjacji charakteryzuje to, że:

Autoenkodery rozwiązujące problem autoasocjacji charakteryzuje to, że:

- ☐ a. rekurencyjnie używają wektora zmiennych ukrytych jako swojego wejścia, by po kilku iteracjach zbliżyć się do finalnego rozwiązania
- ☒ b. oryginalne dane są najpierw reprezentowane jako pewien wektor cech, a następnie są na jego podstawie rekonstruowane ✓
- ☐ c. oryginalne przypadki uczące zostają zapamiętane w wektorze zmiennych ukrytych, a następnie sieć decyduje, który z nich zostanie zwrócony na wyjście
- ☐ d. porównują wektor zmiennych ukrytych oryginalnych danych z wektorem uzyskanym po ich przetworzeniu przez sieć
- ☐ e. do ich uczenia nie można użyć algorytmu wstecznej propagacji błędów

Poprawna odpowiedź to: oryginalne dane są najpierw reprezentowane jako pewien wektor cech, a następnie są na jego podstawie rekonstruowane

Zbyt wysoka wartość współczynnika (prędkości) uczenia może spowodować:

Zbyt wysoka wartość współczynnika (prędkości) uczenia może spowodować:

- ☒ a. pominięcie ("przeskoczenie") optymalnego rozwiązania. ✓
- ☐ b. utknięcie optymalizacji w lokalnym optimum.
- ☐ c. wagi zostaną zaktualizowane w złym kierunku (przeciwnym niż optimum)
- ☐ d. bardzo długi czas uczenia.

Które z poniższych stwierdzeń dot. warstwy poolingu jest nieprawdziwe?

Które z poniższych stwierdzeń dot. warstwy poolingu jest **nieprawdziwe**?

- ☐ a. Jest warstwą nieuczalną.
- ☐ b. Pozwala ograniczyć przetwarzanie podobnych informacji występujących obok siebie w danych.
- ☐ c. Wykonuje pewną funkcję (np. min, max, avg) na wycinku (oknie) danych.
- ☒ d. Spłaszcza dane pozwalając przejść z warstw konwolucyjnych do w pełni połączonych (gęstych) ✓

Dana jest warstwa konwolucyjna zawierająca 64 filtry z oknem (kernel) o rozmiarze 3×3 . Ile wag (zmiennych uczonych) jest w tej warstwie?

Dana jest warstwa konwolucyjna zawierająca 64 filtry z oknem (kernel) o rozmiarze 3×3 . Ile wag (zmiennych uczonych) jest w tej warstwie?

Odpowiedź: ✓

Warstwa flatten:

Warstwa flatten:

- ☒ a. Spłaszcza dane pozwalając przejść z warstw konwolucyjnych do w pełni połączonych (gęstych) ✓
- ☐ b. Jest warstwą uczalną.
- ☐ c. Spłaszcza dane pozwalając przejść z warstw w pełni połączonych (gęstych) do konwolucyjnych
- ☐ d. Wykonuje pewną funkcję (np. min, max, avg) na wycinku (oknie) danych.

Wariacyjny autoenkoder różni się od zwykłego tym, że:

Wariacyjny autoenkoder różni się od zwykłego tym, że

- ☒ a. daje gwarancję, że wektory zmiennych ukrytych są unikalne dla wszystkich przypadków w danych ✗
- ☐ b. uczony jest rozkład prawdopodobieństwa, który wykorzystywany jest do generowania wektora ukrytych zmiennych
- ☐ c. nauczone wektory zmiennych ukrytych rozrzucone są równomiernie po całej przestrzeni
- ☐ d. wariancja wskazań tego autoenkodera dla różnych danych wykorzystywana jest do uczenia sieci

Dany jest zbiór przykładów (obiektów) o atrybutach, których liczbę należy ograniczyć. Wskazać architekturę sztucznej sieci neuronowej, najlepiej nadającą się do tego zadania.

Dany jest zbiór przykładów (obiektów) o atrybutach, których liczbę należy ograniczyć. Wskazać architekturę sztucznej sieci neuronowej, najlepiej nadającą się do tego zadania.

- ☐ a. Sieć wielowarstwowa.
- ☒ b. Sieć typu autoenkoder. ✓
- ☐ c. Sieć konwolucyjna.
- ☐ d. Sieć rekurencyjna.
- ☐ e. Sieć perceptronowa.

Które stwierdzenie dotyczące autoenkodera jest nieprawidłowe?

Które stwierdzenie dotyczące autoenkodera jest **nieprawidłowe**?

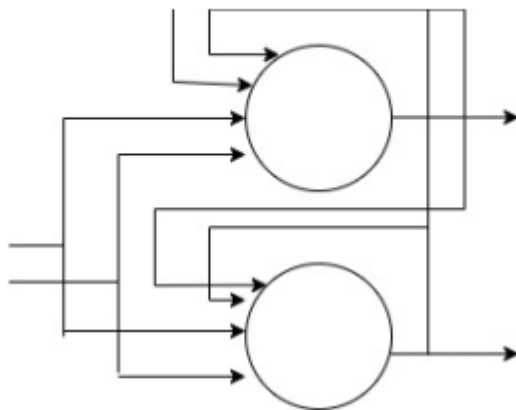
- ☐ a. Autoenkoder składa się z dwóch komponentów: enkodera i dekodera
- ☒ b. Autoenkoder zawiera zawsze dwie warstwy: jedną w module enkodera i jedną w module dekodera ✓
- ☐ c. Autoenkoder może służyć do redukcji wymiarów w danych
- ☐ d. Zadaniem autoenkodera jest otrzymanie na wyjściu informacji tożsamej z wejściem

Która z wymienionych warstw nie jest warstwą regularyzacji?

Która z wymienionych warstw **nie jest** warstwą regularyzacji?

- ☐ a. Weight decay
- ☐ b. Batch normalization
- ☐ c. Dropout
- ☒ d. Flatten ✓

Przedstawiona poniżej warstwa neuronów jest przykładem warstwy:



- ☐ a. konwolucyjnej
- ☐ b. w pełni połączonej
- ☒ c. rekurencyjnej ✓

W algorytmie wstecznej propagacji błędów we wzorze na aktualizację wag: $w'_i = w_i - \mu \frac{\delta L}{\delta w_i}$

W algorytmie wstecznej propagacji błędów we wzorze na aktualizację wag: $w'_i = w_i - \mu \frac{\delta L}{\delta w_i}$

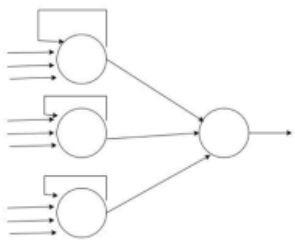
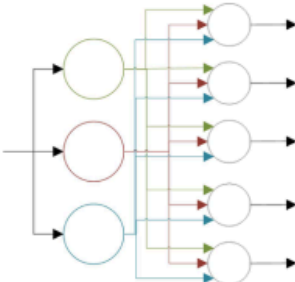
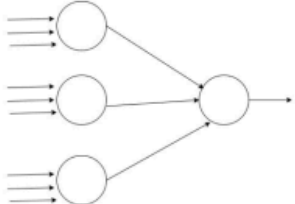
- ☒ a. μ oznacza prędkość uczenia. ✓
- ☐ b. μ oznacza wartość wejściową neuronu dla aktualnego przykładu uczącego.
- ☐ c. μ oznacza wartość funkcji aktywacji neuronu.
- ☐ d. μ oznacza wartość funkcji straty (błędów).

Dany jest zbiór przykładów (obiektów) o atrybutach, których liczbę należy ograniczyć.
Wskazać architekturę sztucznej sieci neuronowej, najlepiej nadającą się do tego zadania.
a. Sieć perceptronowa.

- ☐ b. Sieć rekurencyjna.
- ☒ c. Sieć typu autoenkoder. ✓
- ☐ d. Sieć wielowarstwowa.
- ☐ e. Sieć konwolucyjna.

Która z poniższych architektur sieci neuronowych jest przykładem sieci rekurencyjnej?

Która z poniższych architektur sieci neuronowych jest przykładem sieci rekurencyjnej?

- ☒ a.  ✓
- ☐ b. 
- ☐ c. 

Warstwy konwolucyjne stosuje się najczęściej

Warstwy konwolucyjne stosuje się najczęściej:

- ☐ a. żadna z pozostałych odpowiedzi nie jest prawidłowa
- ☒ b. do przetwarzania danych wielowymiarowych (np. obrazów) ✓
- ☐ c. do analizy przebiegów czasowych
- ☐ d. w prostej klasyfikacji binarnej

Który wzór reprezentuje wyjście zwrócone przez neuron z funkcją aktywacji f , wagami w , biasem b , oraz wejściami x ?

Który wzór reprezentuje wyjście zwrócone przez neuron z funkcją aktywacji f , wagami w_i , biasem b , oraz wejściami x_i ?

Wybierz jedną odpowiedź:

- ☐ a. $\sum_i f(w_i \cdot x_i + b)$
- ☒ b. $f(\sum_i w_i \cdot x_i + b)$ ✓
- ☐ c. $\sum_i w_i \cdot f(b \cdot x_i)$
- ☐ d. $\sum_i w_i \cdot f(x_i + b)$

Jaki typ sieci neuronowych sprawdzi się najlepiej do prognozowania pogody?

Jaki typ sieci neuronowych sprawdzi się najlepiej do prognozowania pogody?

- ☐ a. Sieci warstwowe.
- ☐ b. Sieci typu ResNet.
- ☐ c. Sieci konwolucyjne.
- ☐ d. Sieci autoasocjacyjne.
- ☒ e. Sieci rekurencyjne. ✓

Dana jest sztuczna sieć neuronowa, wykorzystująca wsteczną propagację błędów. W skład sieci wchodzi neuron u , połączony z neuronem v z warstwy następnej, a waga tego połączenia wynosi w . Określić zmianę Δw wagi połączenia w przy założeniu, że sygnał wyjściowy neuronu u ma wartość 0,2, błąd neuronu v wynosi 0,1 a współczynnik uczenia ma wartość 0,5.

Dana jest sztuczna sieć neuronowa, wykorzystująca wsteczną propagację błędów. W skład sieci wchodzi neuron u , połączony z neuronem v z warstwy następnej, a waga tego połączenia wynosi w . Określić zmianę Δw wagi połączenia w przy założeniu, że sygnał wyjściowy neuronu u ma wartość 0,2, błąd neuronu v wynosi 0,1 a współczynnik uczenia ma wartość 0,5.

Odpowiedź: ✗

Poprawna odpowiedź to: 0,1

Wskazać parametry, ze względu na które metoda batch normalization skaluje wyjście warstwy.

Wskazać parametry, ze względu na które metoda *batch normalization* skaluje wyjście warstwy.

- ☒ a. Odchylenie standardowe i średnia arytmetyczna. ✓
- ☐ b. Wartość elementu maksymalnego i wariancja.
- ☐ c. Mediana wartości i średnia arytmetyczna.
- ☐ d. Wariancja i mediana wartości.
- ☐ e. Wartość elementu minimalnego i wariancja.

Twoja odpowiedź jest poprawna.

Poprawna odpowiedź to: Odchylenie standardowe i średnia arytmetyczna.

Dana jest sztuczna sieć neuronowa, wykorzystująca algorytm stochastycznego spadku gradientu (ang. stochastic gradient descent). Wskazać, ile przykładów ze zbioru treningowego należy uwzględnić w każdej iteracji algorytmu, aby zaktualizować wagi połączeń w sieci.

Dana jest sztuczna sieć neuronowa, wykorzystująca algorytm stochastycznego spadku gradientu (ang. *stochastic gradient descent*). Wskazać, ile przykładów ze zbioru treningowego należy uwzględnić w każdej iteracji algorytmu, aby zaktualizować wagi połączeń w sieci.

- ☐ a. Liczba przykładów zależy od parametrów statystycznych zbioru treningowego.
- ☒ b. Dowolną liczbę przykładów, losowaną dla każdej iteracji. ✗
- ☐ c. Dowolną liczbę przykładów, ale równą dla każdej iteracji.
- ☐ d. Jeden przykład.
- ☐ e. Wszystkie przykłady.

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: Jeden przykład.

Wskazać przeznaczenie architektury ResNet (ang. residual neural network)

Wskazać przeznaczenie architektury ResNet (ang. *residual neural network*).

- ☒ a. Przyspieszanie propagacji w przód w sieci konwolucyjnej. ✖
- ☐ b. Rozwiązywanie problemu zanikającego gradientu.
- ☐ c. Normalizacja warstwy wejściowej w sieci konwolucyjnej.
- ☐ d. Sterowanie procesem rozwijania w sieciach rekurencyjnych.
- ☐ e. Zapobieganie przeuczeniu sieci konwolucyjnej.

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: Rozwiązywanie problemu zanikającego gradientu.

Wskazać zadania, do których realizacji zwykle nie używa się tradycyjnych autoenkoderów

Wskazać zadania, do których realizacji zwykle **nie używa** się tradycyjnych autoenkoderów.

- ☒ a. Łączenie kolejnych danych w sekwencje. ✔
- ☐ b. Odszumianie danych.
- ☐ c. Redukowanie wymiarowości danych.
- ☐ d. Grupowanie danych.
- ☐ e. Tworzenie wariantów obrazów.

Dana jest wielowarstwowa, sztuczna sieć neuronowa z liniową funkcją aktywacji w każdej warstwie. Wskazać najprostszą, funkcjonalnie równoważną strukturę, do której taka sieć może być zredukowana.

Dana jest wielowarstwowa, sztuczna sieć neuronowa z liniową funkcją aktywacji w każdej warstwie. Wskazać najprostszą, funkcjonalnie równoważną strukturę, do której taka sieć może być zredukowana.

- ☒ a. Sieć jednowarstwowa z liniową funkcją aktywacji. ✔
- ☐ b. Sieć dwuwarstwowa z liniową funkcją aktywacji.
- ☐ c. Sieć jednowarstwowa z nieliniową funkcją aktywacji.
- ☐ d. Rozpatrywanej sieci nie można zredukować.
- ☐ e. Sieć dwuwarstwowa z nieliniową funkcją aktywacji.

Która z poniższych odpowiedzi dotyczących biasu używanego w neuronie jest prawdziwa?

Która z poniższych odpowiedzi dotyczących biasu używanego w neuronie jest prawdziwa?

Wybierz jedną odpowiedź:

- ☐ a. Bias jest stały podczas uczenia, wagi muszą się dostosować do dobranych wartości biasu.
- ☒ b. Bias jest dodawany przed zaaplikowaniem funkcji aktywacji. ✓
- ☐ c. Neuron podczas agregacji dodaje uśrednioną wartość biasu połączonych z nim neuronów z poprzedniej warstwy.
- ☐ d. Z biasu można zrezygnować bez szkody dla jakości klasyfikacji sieci.

Jakie będą skutki dobrania zbyt dużego współczynnika uczenia (learning rate)?

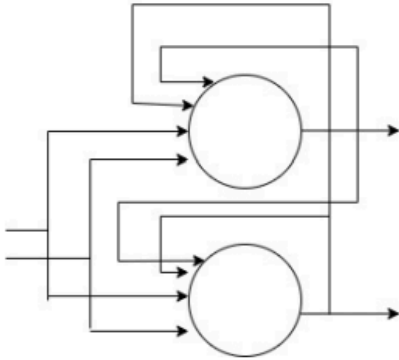
Jakie będą skutki dobrania zbyt dużego współczynnika uczenia (learning rate)?

Wybierz jedną odpowiedź:

- ☐ a. Proces optymalizacji zatrzyma się przedwcześnie z powodu zerowej pochodnej.
- ☐ b. Nauczona sieć będzie działać niedeterministycznie.
- ☒ c. Proces optymalizacji może być niezdolny dotrzeć do wartości bliskich optimum. ✓
- ☐ d. Sieć neuronowa przeuczy się.

Przedstawiona poniżej warstwa neuronów jest przykładem warstwy

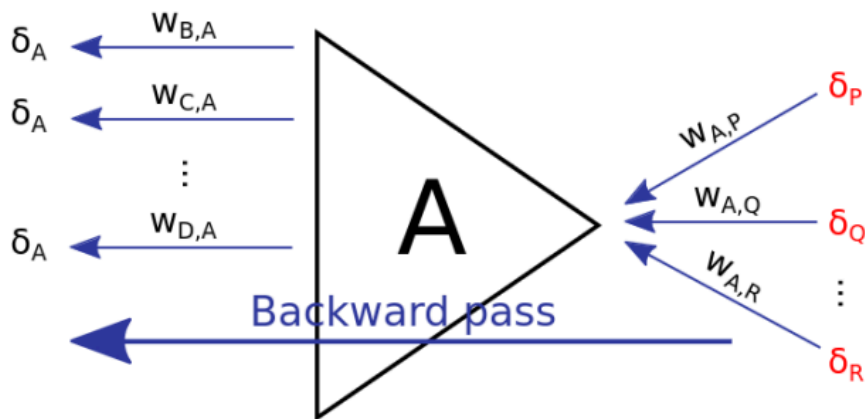
Przedstawiona poniżej warstwa neuronów jest przykładem warstwy:



- ☐ a. autoasocjacyjnej.
- ☐ b. gęstej.
- ☒ c. rekurencyjnej. ✓
- ☐ d. konwolucyjnej.
- ☐ e. w pełni połączonej.

Poniższy obrazek przedstawia zarys działania algorytmu wstecznej propagacji błędów dla pojedynczego neuronu

Poniższy obrazek przedstawia zarys działania algorytmu wstecznej propagacji błędów dla pojedynczego neuronu.



W jaki sposób należy obliczyć zmienną δ_A ? Załóżmy, że $X \in \{P, Q, R\}$

Wybierz jedną odpowiedź:

- ☒ a. Należy obliczyć sumę ważoną wszystkich wchodzących δ_X po wagach $w_{A,X}$, a następnie pomnożyć ją przez pochodną funkcji aktywacji neuronu A, $f'(z_A)$ ✓
- ☐ b. Żadna z pozostałych odpowiedzi nie jest poprawna
- ☐ c. Należy zsumować wszystkie δ_X i pomnożyć sumę przez pochodną funkcji aktywacji neuronu A, $f'(z_A)$
- ☐ d. Należy zsumować wszystkie δ_X
- ☐ e. Należy obliczyć sumę ważoną wszystkich wchodzących δ_X po wagach $w_{A,X}$

Mamy w sieci neuronowej pewną warstwę $W = \text{Dense}(5)$. Następnie postanawiamy dodać do sieci jeszcze jedną w pełni połączoną warstwę $W = \text{Dense}(10)$ po warstwie W . Ile dodatkowych uczalnych parametrów sieci zostanie utworzonych w związku z powiększeniem architektury tej sieci?

Mamy w sieci neuronowej pewną warstwę $W_1 = \text{Dense}(5)$. Następnie postanawiamy dodać do sieci jeszcze jedną w pełni połączoną warstwę $W_2 = \text{Dense}(10)$ po warstwie W_1 . Ile dodatkowych uczalnych parametrów sieci zostanie utworzonych w związku z powiększeniem architektury tej sieci?

Wybierz jedną odpowiedź:

- ☐ a. 35
- ☐ b. 20
- ☐ c. 50
- ☐ d. 55
- ☐ e. 60
- ☒ f. 15 ✗

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: 60

Zadaniem warstwy poolingu jest:

Zadaniem warstwy poolingu jest:

Wybierz jedną odpowiedź:

- ☐ a. Normalizowanie danych
- ☐ b. Agregacja kilku przypadków uczących do jednego tensora
- ☐ c. Zwiększenie wymiarowości danych
- ☒ d. Agregacja wartości z pewnego fragmentu tensora do jednej wartości ✓
- ☐ e. Regularyzacja sieci

Technika wczesnego kończenia uczenia sieci (early stopping) przy wykorzystaniu zbioru walidacyjnego polega na:

Technika wczesnego kończenia uczenia sieci (early stopping) przy wykorzystaniu zbioru walidacyjnego polega na:

Wybierz jedną odpowiedź:

- ☐ a. Usunięciu ze zbioru walidacyjnego wszystkich outlierów.
- ☐ b. Zoptymalizowaniu działania sieci na zbiorze walidacyjnym, a potem dopiero douczeniu jej na zbiorze uczącym.
- ☒ c. Przerwaniu uczenia sieci kiedy błąd na zbiorze walidacyjnym zacznie wzrastać. ✓
- ☐ d. Dodaniu specjalnej warstwy walidacyjnej do sieci neuronowej.

Czym jest autoencoder?

Pytanie **8**

Zakończone

Punkty maks.: 1,00

Czym jest autoencoder?

autoencoder składa się z enkodera i dekodera. Enkoder zmniejsza ilość klas - dekoderek potem te dane dekoduje

Technika regularyzacji batch normalization polega na:

Technika regularyzacji batch normalization polega na:

Wybierz jedną odpowiedź:

- ☐ a. Inicjalizacji wag neuronów zgodnie z rozkładem normalnym $N(0,1)$.
- ☐ b. Standaryzacji wag w neuronach na podstawie aktualnie przetwarzanego batcha przypadków uczących.
- ☐ c. Standaryzacji aktywacji uzyskanych przez warstwę sieci dla przykładów z aktualnie przetwarzanego batcha przypadków uczących.
- ☒ d. Normalizacji przypadków uczących z aktualnie przetwarzanego batcha przypadków uczących. ✖

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to:

Standaryzacji aktywacji uzyskanych przez warstwę sieci dla przykładów z aktualnie przetwarzanego batcha przypadków uczących.

Wybierz prawdziwe zdanie dotyczące padding'u.

Wybierz prawdziwe zdanie dotyczące padding'u.

- ☐ a. Służy do redukcji rozmiaru tensora przy przejściu przez warstwę konwolucyjną.
- ☒ b. Pozwala na zachowanie rozmiarów H i W tensora przy przejściu przez warstwę konwolucyjną. ✔
- ☐ c. Redukuje rozmiar tensora przy przejściu przez warstwę gęstą.
- ☐ d. Pozwala na przekazanie stanu ukrytego w warstwie rekurencyjnej.

Wybierz zdanie prawdziwe dotyczące warstw gęstych.

Wybierz zdanie prawdziwe dotyczące warstw gęstych.

- ☒ a. Każdy neuron z jednej warstwy gęstej jest połączony z każdym neuronem z drugiej warstwy gęstej. ✔
- ☐ b. Ilość połączeń wychodzących od neuronów może ulec zmianie w trakcie propagacji wstecznej.
- ☐ c. Ilość połączeń wychodzących od każdego neuronu sterowana jest hiperparametrem.
- ☐ d. Neurony z sąsiadujących warstw gęstych są łączone parami.

Wybierz zdanie fałszywe dotyczące BatchNormalization.

Wybierz zdanie fałszywe dotyczące BatchNormalization.

- ☐ a. Warstwa ta stabilizuje proces uczenia sieci.
- ☐ b. Warstwa ta przeprowadza standaryzację tensorów wejściowych według statystyk z batch'a.
- ☐ c. Warstwa ta po treningu wykorzystuje średnie statystyki zebrane w trakcie treningu.
- ☒ d. Warstwa ta spowalnia zbieganie się treningu sieci. ✓

Co jest zadaniem autoenkodera w zadaniu autoasocjacji?

Co jest zadaniem autoenkodera w zadaniu autoasocjacji?

Wybierz jedną odpowiedź:

- ☐ a. Wskazanie najważniejszych grup (klastrow) w danych.
- ☒ b. Odtworzenie na wyjściu oryginalnego wejścia. ✓
- ☐ c. Minimalizacja wariancji na wyjściu.
- ☐ d. Klasyfikacja do właściwej klasy decyzyjnej.

Wskaż potencjalną architekturę sieci typu autoenkoder, dokonującą autoasocjacji

Wskaż potencjalną architekturę sieci typu autoenkoder, dokonującą autoasocjacji.

- ☒ a. Dense(100) -> Dense(10) -> Dense (100) ✓
- ☐ b. Dense(50) -> Dense(10) -> Dense (100)
- ☐ c. Dense(20) -> Dense(100) -> Dense (10)
- ☐ d. Dense(100) -> Dense(10) -> Dense (50)

Spośród poniżej przedstawionych architektur, wskaż tę która reprezentuje typowy autoenkoder w zadaniu autoasocjacji.

Spośród poniżej przedstawionych architektur, wskaż tę która reprezentuje typowy autoenkoder w zadaniu autoasocjacji.

Wybierz jedną odpowiedź:

- ☐ a. Dense(100) -> Dense(10) -> Dense(100)
- ☐ b. Dense(100) -> Dense(10) -> Dense(2)
- ☒ c. Dense(100) -> Dense(100) -> Dense(100) ✗
- ☐ d. Dense(100) -> Dense(1000) -> Dense(100)

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to:

Dense(100) -> Dense(10) -> Dense(100)

Dane są dwie warstwy gęste z bias'ami: W1 - 30 neuronów, W2 - 40 neuronów. Ile parametrów uczalnych będzie miała sieć W1->W2, jeżeli na wejście tej sieci wchodzi wektor o długości 30?

Dane są dwie warstwy gęste z bias'ami: W1 - 30 neuronów, W2 - 40 neuronów. Ile parametrów uczalnych będzie miała sieć W1->W2, jeżeli na wejście tej sieci wchodzi wektor o długości 30?

- ☐ a. 1270
- ☒ b. 2170 ✓
- ☐ c. 2000
- ☐ d. 2100

Który wzór jest poprawny dla neuronu z bias'em i wejściem w formie wektora jednoelementowego? y - wyjście neuronu, f - funkcja aktywacji, b - bias, w - waga

Który wzór jest poprawny dla neuronu z bias'em i wejściem w formie wektora jednoelementowego? y - wyjście neuronu, f - funkcja aktywacji, b - bias, w - waga

- ☐ a. $y = f(wx) + b$
- ☐ b. $y = f(w(x + b))$
- ☐ c. $y = f(wb + x)$
- ☒ d. $y = f(wx + b)$ ✓

Wybierz zdanie, które jest prawdziwe

Wybierz zdanie, które jest prawdziwe.

- ☐ a. Wagi w sieciach neuronowych są inicjalizowane zerami.
- ☐ b. Wagi w sieciach neuronowych są inicjalizowane niezerową wartością stałą.
- ☐ c. Wagi w sieciach neuronowych nie wymagają inicjalizacji przed uczeniem.
- ☒ d. Do inicjalizacji wag w sieciach neuronowych używa się przeważnie rozkładu normalnego lub jednostajnego. ✓

Wybierz zdanie prawdziwe odnośnie stochastycznego spadku gradientu.

Wybierz zdanie prawdziwe odnośnie stochastycznego spadku gradientu.

- ☒ a. Przykłady treningowe nie są mieszane. ✗
- ☐ b. Iteracyjnie minimalizowana jest w nim funkcja straty.
- ☐ c. Wykorzystywane są w nim pochodne drugiego rzędu.
- ☐ d. Polega on na minimalizacji wartości wag.

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: Iteracyjnie minimalizowana jest w nim funkcja straty.

Podczas treningu wartość wchodząca na funkcję aktywacji ReLU w trakcie forward pass'a wynosi -1. Ile będzie wynosiła wartość pochodnej funkcji ReLU w backward pass'ie?

Podczas treningu wartość wchodząca na funkcję aktywacji ReLU w trakcie forward pass'a wynosi -1. Ile będzie wynosiła wartość pochodnej funkcji ReLU w backward pass'ie?

- ☐ a. 2
- ☐ b. -1
- ☒ c. 0 ✓
- ☐ d. 1

Wybierz zdanie fałszywe dla warstwy Dropout.

Wybierz zdanie fałszywe dla warstwy Dropout.

- ☐ a. Zeruje pewną ilość wartości w tensorze wejściowym.
- ☐ b. Przeciwdziała przeuczeniu.
- ☐ c. Przy złej parametryzacji może uniemożliwić trening.
- ☒ d. Funkcjonuje w treningu sieci i w trakcie odpytywania już nauczonej sieci. ✓

Waga wynosi 0.5, pochodna cząstkowa funkcji kosztu po tej wadze wynosi 0.4, a współczynnik uczenia (learning rate) wynosi 0.01. Jaka będzie wartość wagi po jej aktualizacji?

Waga wynosi 0.5, pochodna cząstkowa funkcji kosztu po tej wadze wynosi 0.4, a współczynnik uczenia (learning rate) wynosi 0.01. Jaka będzie wartość wagi po jej aktualizacji?

- ☐ a. 0.54
- ☐ b. 0.504
- ☐ c. 0.46
- ☒ d. 0.496 ✓

Mamy warstwę konwolucyjną o liczbie filtrów równej 32, kroku równym 1 i rozmiarze kernela równym 3x3. Jaka będzie liczba kanałów (ilość map cech) w tensorze wyjściowym z tej warstwy?

Mamy warstwę konwolucyjną o liczbie filtrów równej 32, kroku równym 1 i rozmiarze kernela równym 3x3. Jaka będzie liczba kanałów (ilość map cech) w tensorze wyjściowym z tej warstwy?

- ☐ a. 64
- ☐ b. 96
- ☒ c. 32 ✓
- ☐ d. 128

Wybierz zdanie prawdziwe dotyczące sieci Resnet

Wybierz zdanie prawdziwe dotyczące sieci Resnet.

- ☐ a. Wykorzystuje ona warstwy rekurencyjne.
- ☐ b. Wykorzystuje ona MinPooling.
- ☒ c. Wykorzystuje ona GlobalAveragePooling. ✓
- ☐ d. Wykorzystuje ona MaxPooling.

Wybierz zdanie prawdziwe dotyczące sieci Resnet

Wybierz zdanie prawdziwe dotyczące sieci Resnet.

- ☐ a. Jest to sieć rekurencyjna
- ☐ b. Sieć składa się z bloków, które posiadają kilka warstw konwolucyjnych pod rząd, zakończonych MaxPooling'iem.
- ☒ c. Wykorzystanie skip connections niweluje problem spadku wyników bardzo głębokich sieci w treningu. ✓
- ☐ d. W tej architekturze stosuje się 3 warstwy gęste na końcu sieci.

W algorytmie wstecznej propagacji błędu:

W algorytmie wstecznej propagacji błędu:

- ☐ a. Żadna z pozostałych odpowiedzi nie jest poprawna.
- ☐ b. Wagi jednostek aktualizuje się zaczynając od pierwszej warstwy.
- ☐ c. Zmiana (aktualizacja) wag wszystkich jednostek jest jednakowa.
- ☒ d. Wagi jednostek aktualizuje się zaczynając od ostatniej warstwy. ✓

Algorytm spadku gradientu (klasyczny):

Algorytm spadku gradientu (klasyczny):

- ☐ a. jest tym samym co algorytm stochastycznego spadku gradientu
- ☐ b. pozwala "uciec" z lokalnego optimum
- ☒ c. wyznacza aktualizacje wag w oparciu o wszystkie przykłady uczące ✓
- ☐ d. jest szybszy od algorytmu stochastycznego spadku gradientu

Czym jest regularyzacja w sieciach neuronowych?

Czym jest regularyzacja w sieciach neuronowych?

Jaki jest cel stosowania regularyzacji z sieciach neuronowych?

Pytanie **11**

Zakończone

Punkty: 1,00 z 1,00

Jaki jest cel stosowania regularyzacji w sieciach neuronowych?

Regularyzację stosuje się w celu uniknięcia efektu przeuczenia

Komentarz:

Czym jest autoencoder?

Czym jest autoencoder?

Tekst odpowiedzi Pytanie 5

Autoenkoder składa się z dwóch komponentów: enkodera i dekodera, jego zadaniem jest otrzymanie na wyjściu informacji tożsamej z wejściem, zawiera zawsze dwie warstwy: jedną w module enkodera i jedną w module dekodera.


Komentarz:

Enkoder i dekoderek nie musi zawierać tylko po jednej warstwie.

MEISSENERA

Wskazać zdanie fałszywe w odniesieniu do architektury LSTM (ang. long short-term memory).

Wskazać zdanie **fałszywe** w odniesieniu do architektury LSTM (ang. *long short-term memory*).

- ☐ a. Następny stan komórki LSTM może zależeć od stanu bieżącego.
- ☐ b. Bieżący stan komórki LSTM zawsze zależy od stanu poprzedniego.
- ☐ c. Architektura LSTM może być wielowarstwowa.
- ☒ d. W architekturze LSTM może zachodzić zjawisko zanikania gradientu. 
- ☐ e. Komórka LSTM zawiera bramki sterujące procesem zapamiętywania.

Twoja odpowiedź jest niepoprawna.


Poprawna odpowiedź to: Bieżący stan komórki LSTM zawsze zależy od stanu poprzedniego.



Dany jest sztuczny neuron z wektorem wag $w=(1.0, 1.0, 0.5)$, biasem $b=0.5$ i z sigmoidalną funkcją aktywacji. Wskazać wartość sygnału wyjściowego, jeżeli na wejście zostanie podany wektor $x=(-2.0, 1.0,$

1.0).

Dany jest sztuczny neuron z wektorem wag $\mathbf{w}=(1.0, 1.0, 0.5)$, biasem $b=0.5$ i z sigmoidalną funkcją aktywacji. Wskazać wartość sygnału wyjściowego, jeżeli na wejście zostanie podany wektor $\mathbf{x}=(-2.0, 1.0, 1.0)$.


- ☐ a. -1.0
- ☒ b. 0.0 
- ☐ c. 1.0
- ☐ d. 0.5
- ☐ e. -0.5

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: 0.5

Regularyzacja typu dropout ma na celu:

Regularyzacja typu *dropout* ma na celu:

- ☐ a. eliminowanie przykładów odstających ze zbioru treningowego.
- ☐ b. spowodowanie, aby jak najwięcej neuronów uczestniczyło w procesie uczenia.
- ☒ c. usuwanie z sieci neuronów nadmiarowych. 
- ☐ d. redukcję wymiaru wektora wejściowego.
- ☐ e. scalanie warstw sieci.

Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: spowodowanie, aby jak najwięcej neuronów uczestniczyło w procesie uczenia.

Dana jest dwuwarstwowa, sztuczna sieć neuronowa z logistyczną funkcją aktywacji w każdej warstwie. Wskazać najogólniejszą

klasę funkcji, które taki model może aproksymować z dowolną dokładnością.

Dana jest dwuwarstwowa, sztuczna sieć neuronowa z logistyczną funkcją aktywacji w każdej warstwie. Wskazać najogólniejszą klasę funkcji, które taki model może aproksymować z dowolną dokładnością.

- ☒ a. Funkcje algebraiczne. ❌
- ☐ b. Funkcje wielomianowe.
- ☐ c. Funkcje ciągłe.
- ☐ d. Funkcje przestępne.
- ☐ e. Funkcje liniowe.

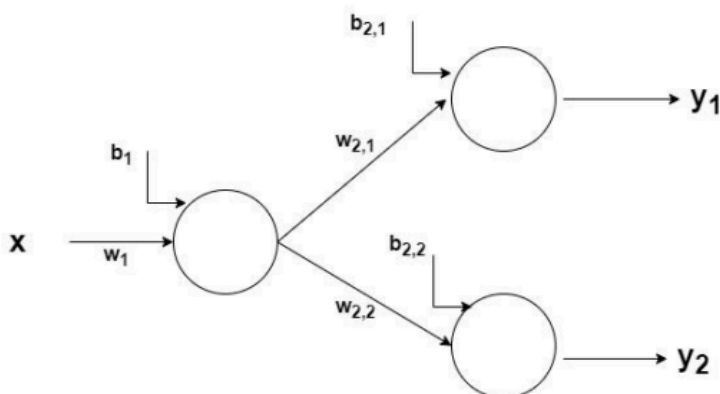
Twoja odpowiedź jest niepoprawna.

Poprawna odpowiedź to: Funkcje ciągłe

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości δ w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: $\delta_{2,1} = 1$, $\delta_{2,2} = 1$. Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące: $x = 2$, $w_1 = 2$, $b_1 = 0$, $w_{2,1} = 4$, $w_{2,2} = 0.5$, $b_{2,1} = 1$, $b_{2,2} = 5$, a współczynnik uczenia (μ) wynosi 0.1

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości δ w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: $\delta_{2,1} = 1$, $\delta_{2,2} = 1$. Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące:

$x = 2$, $w_1 = 2$, $b_1 = 0$, $w_{2,1} = 4$, $w_{2,2} = 0.5$, $b_{2,1} = 1$, $b_{2,2} = 5$ a współczynnik uczenia (μ) wynosi 0.1.



Odpowiedź: 1.9

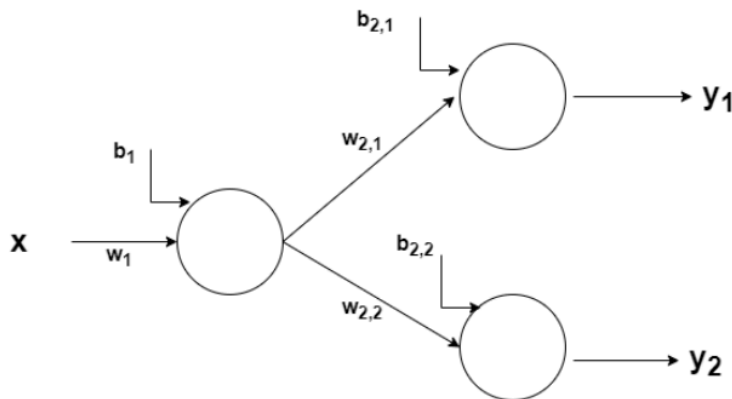


Poprawna odpowiedź to: 1,1

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości w algorytmie wstecznej

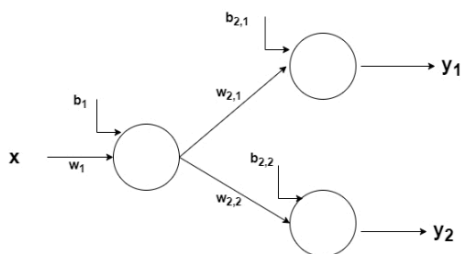
propagacji błędu dla jednostek drugiej warstwy są następujące: , . Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące: a współczynnik uczenia (μ) wynosi 0.1.

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości δ w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: $\delta_{2,1} = 0$, $\delta_{2,2} = 1$. Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące: $x = 1$, $w_1 = 0.5$, $b_1 = 0$, $w_{2,1} = 0.1$, $w_{2,2} = 2$, $b_{2,1} = 1$, $b_{2,2} = 5$ a współczynnik uczenia (μ) wynosi 0.1.



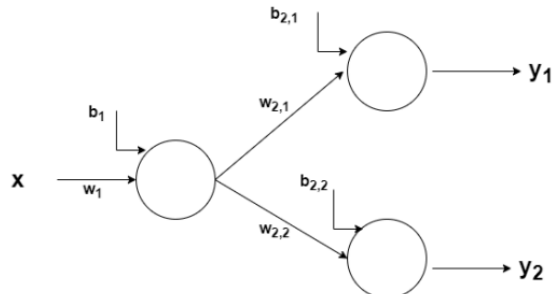
Odpowiedź:

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości δ w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: $\delta_{2,1} = 1$, $\delta_{2,2} = 1$. Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące: $x = 2$, $w_1 = 2$, $b_1 = 0$, $w_{2,1} = 4$, $w_{2,2} = 0.5$, $b_{2,1} = 1$, $b_{2,2} = 5$ a współczynnik uczenia (μ) wynosi 0.1.



Odpowiedź:

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości δ w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: $\delta_{2,1} = 1$, $\delta_{2,2} = 2$. Jaka będzie wartość wagi w_1 po aktualizacji, jeśli wartości przed aktualizacją są następujące: $x = 6$, $w_1 = 0$, $b_1 = 0$, $w_{2,1} = -0.5$, $w_{2,2} = 0.5$, $b_{2,1} = 1$, $b_{2,2} = 5$ a współczynnik uczenia (μ) wynosi 0.1.



Odpowiedź: ✔

Dana jest sieć neuronowa przedstawiona na poniższym schemacie. We wszystkich jednostkach stosuje się liniową funkcję aktywacji. Wartości w algorytmie wstecznej propagacji błędów dla jednostek drugiej warstwy są następujące: , . Jaka będzie wartość wagi po aktualizacji, jeśli wartości przed aktualizacją są następujące: a współczynnik uczenia () wynosi 0.1.

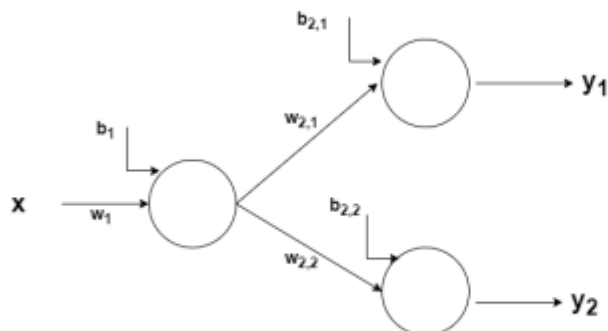
Odpowiedź: -0,4

δ

$\delta_{2,1} = 2$ $\delta_{2,2} = 0$ w_1

$x = 5$, $w_1 = 0.1$, $b_1 = 0$, $w_{2,1} = 0.5$, $w_{2,2} = 0.5$, $b_{2,1} = 1$, $b_{2,2} = 5$ μ

wynosi 0.1.



Odpowiedź: ✔

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przypadku X_1 przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.

Schemat sieci neuronowej:

Funkcja aktywacji: ReLU

Schemat sieci neuronowej.



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X_1 = 4$, $Y_1 = 0$

Odpowiedź powinna zawierać:

• Obliczenia dla:

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

• Wzór na:

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$$z(1) = w(1) * X + b = -1,5 * 4 + 0,5 = -6 + 0,5 = -5,5$$

$$y(1) = \text{ReLU}(-5,5) = 0$$

$$z(2) = w(2) * y(1) + b = 1 * 0 - 0,5 = -0,5$$

$$y(2) = \text{ReLU}(-0,5) = 0$$

$$\text{Delta}(2) = 2 * (0 - 0) * \text{Act}'(z2) = 2 * 0 * 0 = 0$$

$$\text{Delta}(1) = \text{Delta}(2) * w(2) * \text{Act}'(z1) = 0 * 1 * 0 = 0$$

u - oznacza stałą uczenia

$$w(1)' = w(1) - u * \text{Delta}(1) * X = -1,5 - (0,1 * 0 * 4) = -1,5$$

$$b(1)' = b(1) - u * \text{Delta}(1) = 0,5 - (0,1 * 0) = 0,5$$

$$w(2)' = w(2) - u * \text{Delta}(2) * y(1) = 1 - 0,1 * 0 * 0 = 1$$

$$b(2)' = b(2) - u * \text{Delta}(2) = -0,5 - 0,1 * 0 = -0,5$$

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przypadku X1 przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.

Schemat sieci neuronowej:



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: X1 = 4, Y1 = 0

Odpowiedź powinna zawierać:

• **Obliczenia dla:**

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

• **Wzór na:**

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

Dane: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$, $lr = 0.1$

Wzory:

$\text{ReLU} = \max(0, z)$

$\text{MSE}: E = 1/2 * (y(2) - Y)^2$

$w(2)' = w(2) - lr * E' * \text{ReLU}'(z(2)) * y(1) = w(2) - lr * (y(2) - Y) * \text{ReLU}'(z(2)) * y(1)$

$b(2)' = b(2) - lr * E' * \text{ReLU}'(z(2)) = w(2) - lr * (y(2) - Y) * \text{ReLU}'(z(2))$

$w(1)' = w(1) - lr * E' * \text{ReLU}'(z(2)) * w(2) * \text{ReLU}'(z(1)) * X = w(1) - lr * (y(2) - Y) * \text{ReLU}'(z(2)) * w(2) * \text{ReLU}'(z(1)) * X$

$b(1)' = b(1) - lr * E' * \text{ReLU}'(z(2)) * w(2) * \text{ReLU}'(z(1)) = w(1) - lr * (y(2) - Y) * \text{ReLU}'(z(2)) * w(2) * \text{ReLU}'(z(1))$

Obliczenia:

$z(1) = 4 * (-1.5) + 0.5 = -5.5$

$y(1) = \text{ReLU}(-5.5) = 0$

$z(2) = 0 * 1 - 0.5 = -0.5$

$y(2) = \text{ReLU}(-0.5) = 0$

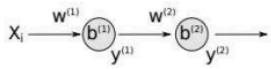
$w(2)' = 1 - 0.1 * (0 - 0) * 0 * 0 = 1$

$b(2)' = -0.5 - 0.1 * (0 - 0) * 0 = -0.5$

$w(1)' = -1.5 - (0 - 0) * 0 * 1 * 0 * 1 = -1.5$

$b(1)' = 0.5 - (0 - 0) * 0 * 1 * 0 = 0.5$

Schemat sieci neuronowej:



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X1 = 4$, $Y1 = 0$

Odpowiedź powinna zawierać:

- Obliczenia dla:

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

- Wzór na:

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$w(1) = -1.5$, $b(1) = 0.5$,

$w(2) = 1$, $b(2) = -0.5$.

$X1 = 4$

$Y1 = 0$

$z1 = x1 * w(1) + b1 = 4 * (-1.5) + 0.5 = -6 + 0.5 = -5.5$

$y1 = \text{ReLU}(z1) = \max(0, z1) = 0$

$z2 = y1 * w(2) + b2 = 0 * 1 - 0.5 = -0.5$

$y2 = \text{ReLU}(z2) = \max(0, z2) = 0$

$w(1)' = w(1) - \mu * dE/dw(1)$

$dE/dw(1) = dE/dy(2) * dy(2)/dz(2) * dz(2)/dy(1) * dy(1)/dz(1) * dz(1)/dw(1) =$

$2(Y1 - y(2)) * (1 \text{ dla } z2 > 0 \text{ i } 0 \text{ dla } z2 \leq 0) * w(2) * (1 \text{ dla } z1 > 0 \text{ i } 0 \text{ dla } z1 \leq 0) * x1 =$

$2 * (0 - 0) * 0 * 1 * 0 * 4 = 0$

$w(1)' = -1.5 - 0.1 * 0 = -1.5$

$b(1)' = b(1) - \mu * dE/db(1)$

$dE/db(1) = dE/dy(2) * dy(2)/dz(2) * dz(2)/dy(1) * dy(1)/dz(1) * dz(1)/db(1) =$

$2(Y1 - y(2)) * (1 \text{ dla } z2 > 0 \text{ i } 0 \text{ dla } z2 \leq 0) * w(2) * (1 \text{ dla } z1 > 0 \text{ i } 0 \text{ dla } z1 \leq 0) * 1 =$

$2 * (0 - 0) * 0 * 1 * 0 * 1 = 0$

$b(1)' = 0.5 - 0.1 * 0 = 0.5$

$b(2)' = b(2) - \mu * dE/db(2)$

$dE/db(2) = dE/dy(2) * dy(2)/dz(2) * dz(2)/db(2) =$

$2(Y1 - y(2)) * (1 \text{ dla } z2 > 0 \text{ i } 0 \text{ dla } z2 \leq 0) * 1 =$

$2 * (0 - 0) * 0 * 1 = 0$

$b(2)' = -0.5 - 0.1 * 0 = -0.5$



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X1 = 4, Y1 = 0$

Odpowiedź powinna zawierać:

• **Obliczenia dla:**

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

• **Wzór na:**

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$$z1 = X * w(1) + b(1) = 4 * (-1.5) + 0.5 = -5.5$$

$$y1 = 0 \text{ (ReLU z minusa = 0)}$$

$$z2 = y1 * w(2) + b(2) = 0 * 1 - 0.5 = -0.5$$

$$y2 = 0 \text{ (ReLU z minusa = 0)}$$

$$\delta 2 = 2(y2 - Y) \text{ (pochodna z } (y2 - Y)^2) = 2 * y2 = 0$$

$$\delta 1 = \delta 2 * w(2) = \delta 2 * 1 = \delta 2 = 0$$

$$w(1)' = w(1) - u * X(\delta 1) = -1.5 - 0.1 * 4 * (\delta 1) = -1.5$$

$$b(1)' = b(1) - u(\delta 1) = 0.5 - 0.1 * (\delta 1) = 0.5$$

$$w(2)' = w(2) - u * y1 * (\delta 2) = 1 - 0.1 * 0 * (\delta 2) = 1$$

$$b(2)' = b(2) - u(\delta 2) = -0.5 - 0.1 * (\delta 2) = -0.5$$

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przypadku X_1 przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X_1 = 4$, $Y_1 = 0$

Odpowiedź powinna zawierać:

• **Obliczenia dla:**

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

• **Wzór na:**

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$$z(1) = 4 * (-1.5) + 0.5 = -6 + 0.5 = -5.5 \quad (\text{wzór: } X * w(1) + b(1))$$

$$y(1) = \text{ReLU}(-5.5) = 0 \quad (\text{wzór: } \text{ReLU}(z(1)))$$

$$z(2) = 0 * 1 + (-0.5) = -0.5 \quad (\text{wzór: } y(1) * w(2) + b(2))$$

$$y(2) = \text{ReLU}(-0.5) = 0 \quad (\text{wzór: } \text{ReLU}(z(2)))$$

Z racji, że błędy wynoszą 0 oraz same pochodne wartości $y(1)$ oraz $y(2)$ również wynoszą 0, to te wartości pozostaną poniżej bez zmian:

$$w(1)' = -1.5$$

$$b(1)' = 0.5$$

$$w(2)' = 1$$

$$b(2)' = -0.5$$

$$\text{Wzór na } w(1)': w(1) - \mu * (\text{pochodna cząstkowa}(E) / \text{pochodna cząstkowa}(w(1)))$$

$$\text{Wzór na } b(1)': b(1) - \mu * (\text{pochodna cząstkowa}(E) / \text{pochodna cząstkowa}(b(1)))$$

$$\text{Wzór na } w(2)': w(2) - \mu * (\text{pochodna cząstkowa}(E) / \text{pochodna cząstkowa}(w(2)))$$

$$\text{Wzór na } b(2)': b(2) - \mu * (\text{pochodna cząstkowa}(E) / \text{pochodna cząstkowa}(b(2)))$$

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.

Schemat sieci neuronowej:



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X_1 = 4$, $Y_1 = 0$

Odpowiedź powinna zawierać:

- **Obliczenia dla:**

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

- **Wzór na:**

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$$z(1) = x_1 * w(1) + b(1) = 4 * (-1,5) + 0,5 = -6 + 0,5 = -5,5$$

$$y(1) = \text{ReLU}(z1) = \max(0, z1) = \max(0, -5,5) = 0$$

$$z(2) = y1 * w2 + b2 = 0 + 0,5 = 0,5$$

$$y(2) = \text{ReLU}(z2) = \max(0, 0,5) = 0,5$$

WZORY i obliczenia dla $w1$, $b1$, $w2$, $b2$: (reszta obliczeń na kartce)

$$w(1)' = w(1) - \mu * \text{pochodna}(E/w1) = -1,5 - 0,1 * 0 = -1,5$$

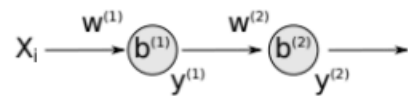
$$b(1)' = b(1) - \mu * \text{pochodna}(E/b1) = 0,5 - 0,1 * 0 = 0,5$$

$$w(2)' = w(2) - \mu * \text{pochodna}(E/w2) = 1 - 0,1 * 0 = 1$$

$$b(2)' = b(2) - \mu * \text{pochodna}(E/b2) = -0,5 - 0,1 * (-1) = -0,4$$

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przypadku X_1 przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.

Schemat sieci neuronowej:



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X_1 = 4$, $Y_1 = 0$

Odpowiedź powinna zawierać:

- **Obliczenia dla:**
 - $z(1) =$

- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

- **Wzór na:**

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

- **Obliczenia dla:**

- $z(1) = w_1 * x + b_1 = -1.5 * 4 + 0.5 = -5.5$
- $y(1) = \text{ReLU}(z_1) = \text{ReLU}(-5.5) = 0$
- $z(2) = w_2 * y_1 + b_2 = 1 * 0 + -0.5 = -0.5$
- $y(2) = \text{ReLU}(z_2) = \text{ReLU}(-0.5) = 0$
- $w(1)' = w_1 - y_1 * u = -1.5$
- $b(1)' = 0.5$
- $w(2)' = w_2 - y_2 * u = 1$
- $b(2)' = -0.5$

- **Wzór na:**

- $w(1)' = w_1 - y_1 * u$
 - $b(1)'$
 - $w(2)' = w_2 - y_2 * u$
 - $b(2)'$
-

Korzystając z algorytmu spadku wzdłuż gradientu, oblicz zaktualizowane wartości wag i biasów dla przypadku X1 przy stałej uczenia $\mu = 0.1$ oraz następujących wartościach parametrów sieci: $w(1) = -1.5$, $b(1) = 0.5$, $w(2) = 1$, $b(2) = -0.5$.

Schemat sieci neuronowej:



Funkcja aktywacji: ReLU

Funkcja błędu: MSE

Zbiór uczący: $X1 = 4$, $Y1 = 0$

Odpowiedź powinna zawierać:

• Obliczenia dla:

- $z(1) =$
- $y(1) =$
- $z(2) =$
- $y(2) =$
- $w(1)' =$
- $b(1)' =$
- $w(2)' =$
- $b(2)' =$

• Wzór na:

- $w(1)'$
- $b(1)'$
- $w(2)'$
- $b(2)'$

Tekst odpowiedzi Pytanie 19

$$z(1) = -1,5 * 4 + 0,5 = -5,5$$

$$y(1) = \max(0, z(1)) = \max(0, -5,5) = 0$$

$$z(2) = y(1) * w(1) + b(1) = 0 * 1 - 0,5 = -0,5$$

$$y(2) = \max(0, z(2)) = 0$$

pochodna z funkcji aktywacji ReLU dla $z(1) = -5,5$ wynosi 0.

$$\delta(2) = 0 * 0 = 0$$

$$w(2)' = 1 - 0,1 * 0 * 0 = 1$$

$$b(2)' = -0,5 - 0,1 * 0 = -0,5$$

$$\delta(1) = 0 * 1 * 0 = 0$$

$$w(1)' = -1,5 - 0,1 * 0 * 4 = -1,5$$

$$b(1)' = 0,5 - 0,1 * 0 = 0,5$$

wzory:

$$\delta(1) = \delta(2) * w(2) * \text{Act}'(z(1))$$

$$\delta(2) = \text{pochodna z błędu MSE po } y(2) * \text{Act}'(z(2))$$

$$w(1)' = w(1) - \mu * X1 * \delta(1)$$

$$b(1)' = b(1) - \mu * \delta(1)$$

STARE JAKIES 2024