



POLITECHNIKA
LUBELSKA
WYDZIAŁ PODSTAW
TECHNIKI

KIERUNEK: MATEMATYKA



Praca magisterska

Przegląd autoenkoderów stosowanych w nienadzorowanym uczeniu
maszynowym

An overview of autoencoders used in unsupervised machine learning

Praca wykonana pod kierunkiem:
dra Dariusza Majerka

Autor:
Alicja Hołowiecka
nr albumu: 89892

Lublin 2022

Spis treści

Wstęp	5
Rozdział 1. Przegląd autoenkoderów	7
1.1. Sztuczne sieci neuronowe	7
1.2. Sieci splotowe	11
1.3. Czym jest autoenkoder	11
Rozdział 2. Przykłady zastosowań autoenkoderów	13
2.1. Analiza PCA za pomocą autoenkodera niedopełnionego	13
2.2. Kolejny przykład	16
Podsumowanie i wnioski	17
Bibliografia	19
Spis rysunków	21
Spis tabel	23
Załączniki	25
Streszczenie (Summary)	27

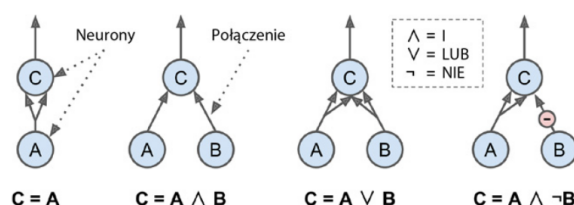
Wstęp

Rozdział 1

Przegląd autoenkoderów

1.1. Sztuczne sieci neuronowe

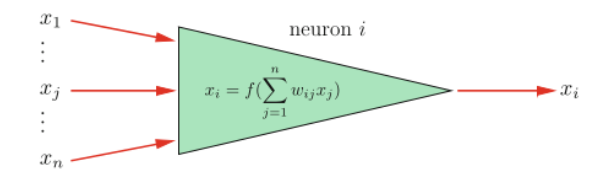
Sztuczny neuron ma co najmniej jedno binarne wejście i dokładnie jedno binarne wyjście. Wyjście jest uaktywniane, jeżeli jest aktywna określona liczba wejść. Na rysunku 1.1 przedstawione są przykładowe sztuczne sieci neuronowe (SSN lub z angielskiego ANN) wykonujące różne operacje logiczne, przy założeniu, że neuron uaktywni się, gdy przynajmniej dwa wejścia będą aktywne [6].



Rysunek 1.1: Przykładowe sztuczne sieci neuronowe rozwiązujące proste zadania logiczne

Źródło: [6]

Jedną z najprostszych architektur SSN jest **perceptron**, którego podstawą jest sztuczny neuron zwany **progową jednostką logiczną** (ang. *Threshold Logic Unit* - TLU) lub **liniową jednostką progową** (ang. *Linear Threshold Unit* - LTU). Wartościami wejść i wyjść są liczby, a każde połączenie ma przyporządkowaną wagę. Jednostka TLU oblicza ważoną sumę sygnałów wejściowych, a następnie zostaje użyta funkcja skokowa. Schemat takiej jednostki został przedstawiony na rysunku 1.2.



Rysunek 1.2: Struktura sztucznego neuronu, który stosuje funkcję skokową f na ważonej sumie sygnałów wejściowych

Źródło: [5]

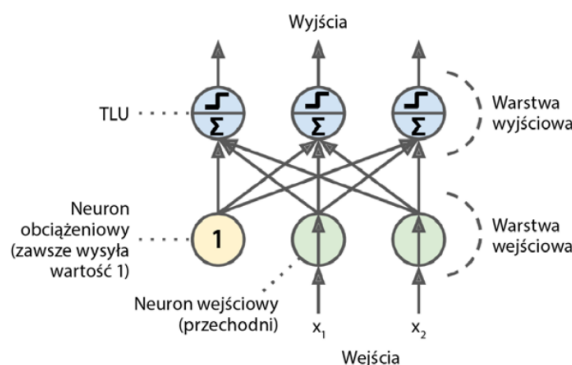
Często używaną funkcją skokową jest **funkcja Heaviside'a**, określona równaniem

$$H(z) = \begin{cases} 0, & \text{jeśli } z < 0 \\ 1, & \text{jeśli } z \geq 0 \end{cases}$$

Czasami stosuje się również **funkcję signum**

$$\text{sgn}(z) = \begin{cases} -1 & \text{jeśli } z < 0 \\ 0, & \text{jeśli } z = 0 \\ 1, & \text{jeśli } z > 0 \end{cases}$$

Perceptron jest złożony z jednej warstwy jednostek TLU, w której każdy neuron jest połączony ze wszystkimi wejściami. Tego typu warstwa jest nazywana **warstwą gęstą**. Warstwa, do której są dostarczane dane wejściowe, jest nazywana **warstwą wejściową** (ang. *input layer*). Najczęściej do tej warstwy jest wstawiany również **neuron obciążeniowy** (ang. *bias neuron*) $x_0 = 1$, który zawsze wysyła wartość 1. Na rysunku 1.3 znajduje się perceptron z dwoma neuronami wejściowymi i jednym obciążeniowym, a także z trzema neuronami w warstwie wyjściowej.



Rysunek 1.3: Perceptron z trzema neuronami wejściowymi i trzema wyjściami

Źródło: [6]

Obliczanie sygnałów wyjściowych w warstwie gęstej przedstawia się wzorem

$$h_{\mathbf{W}, \mathbf{b}}(\mathbf{X}) = \phi(\mathbf{XW} + \mathbf{b})$$

gdzie \mathbf{X} - macierz cech wejściowych, \mathbf{W} - macierz wag połączeń (oprócz neuronu obciążeniowego), \mathbf{b} - wektor obciążeń zawierający wagi połączeń neuronu obciążeniowego ze wszystkimi innymi neuronami, ϕ - tzw. **funkcja aktywacji**, w przypadku TLU jest to funkcja skokowa.

Algorytm uczący, który służy do trenowania perceptronu, jest silnie inspirowany działaniem neuronu biologicznego. Gdy biologiczny neuron często pobudza inną komórkę nerwową, to połączenia między nimi stają się silniejsze. Reguła ta jest nazywana **regułą Hebba**.

Perceptrony są uczone za pomocą odmiany tej reguły, w której połączenia są wzmacniane, jeśli pomagają zmniejszyć wartość błędu. Dokładniej, w danym momencie perceptron przetwarza jeden przykład uczący i wylicza dla niego predykcję. Na każdy neuron wyjściowy odpowiadający za nieprawidłową prognozę następuje zwiększenie wag połączeń ze wszystkimi wejściami przyczyniającymi się do właściwej prognozy. Aktualizowanie wag przedstawia się następującym wzorem

$$\Delta w_{ij} = \eta(y_j - \hat{y}_j)x_i$$

gdzie w_{ij} - waga połączenia między i -tym neuronem wejściowym a j -tym neuronem wyjściowym, x_i - i -ta wartość wejściowa bieżącego przykładu uczącego, \hat{y}_j - wynik j -tego neuronu wyjściowego dla bieżącego przykładu uczącego, y_j - docelowy wynik j -tego neuronu, η - współczynnik uczenia.

Perceptron ma wiele wad związanych z niemożnością rozwiązania pewnych trywialnych problemów (np. zadanie klasyfikacji rozłącznej czyli XOR). Część tych ograniczeń można wyeliminować, stosując architekturę SSN złożoną z wielu warstw perceptronów, czyli **perceptron wielowarstwowy** (ang. *Multi-Layer Perceptron*). Składa się on z jednej warstwy wejściowej (przechodniej), co najmniej jednej warstwy jednostek TLU - tzw. **warstwy ukryte** (ang. *latent layers*) i ostatniej warstwy jednostek TLU - warstwy wyjściowej. Oprócz warstwy wejściowej każda warstwa zawiera neuron obciążający i jest w pełni połączona z następną warstwą. Sieć zawierająca wiele warstw ukrytych nazywamy **głębką siecią neuronową** (ang. *Deep Neural Network* - DNN).

Do uczenia perceptronów wielowarstwowych wykorzystywany jest algorytm **propagacji wstecznej** (ang. *backpropagation*). Propagacja wsteczna jest właściwie algorytmem gradientu prostego [9]. Można go zapisać jako

$$w_{updated} = w_{old} - \eta \nabla E$$

gdzie E jest funkcją kosztu (funkcją straty) [9]. Proces jest powtarzany do momentu uzyskania zbieżności z rozwiązaniem, a każdy przebieg jest nazywany **epoką** (ang. *epoch*).

Uwaga 1.1. Wagi połączeń wszystkich warstw ukrytych należy koniecznie zainicjować losowo. W przeciwnym przypadku proces uczenia zakończy się niepowodzeniem. Na przykład jeśli wszystkie wagi i obciążenia zostaną zainicjowane wartością 0, to model będzie działał tak, jak gdyby składał się tylko z jednego neuronu. Przy zainicjowaniu wag losowo, symetria zostanie złamana i algorytm propagacji wstecznej będzie w stanie wytrenować zespół zróżnicowanych neuronów [6].

Aby algorytm propagacji wstecznej działał prawidłowo, kluczową zmianą jest zastąpienie funkcji skokowej przez inne **funkcje aktywacji**. Zmiana ta jest konieczna, ponieważ funkcja skokowa zawiera jedynie płaskie segmenty i przez to nie pozwala korzystać z gradientu.

Najczęściej używana jest **funkcja logistyczna (sigmoidalna)**

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Ma ona w każdym punkcie zdefiniowaną pochodną niezerową, dzięki czemu algorytm gradientu prostego może na każdym etapie uzyskać lepsze wyniki. Zbiór wartości tej funkcji wynosi od 0 do 1.

Inną popularną funkcją aktywacji jest **tangens hiperboliczny**

$$\tanh(z) = 2\sigma(2z) - 1$$

Funkcja ta jest ciągła i różniczkowalna, a jej zakres wartości wynosi -1 do 1 . Dzięki temu zakresowi wartości wynik każdej warstwy jest wyśrodkowany wobec zera na początku uczenia, co często pomaga w szybszym uzyskaniu zbieżności.

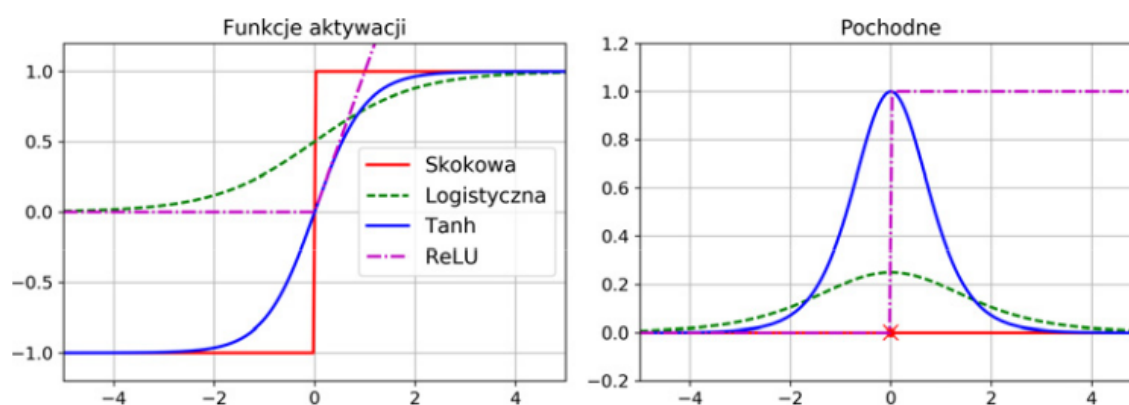
Wśród popularnych funkcji aktywacji należy także wyróżnić **funkcję ReLU** (ang. *Rectified Linear Unit* - prostowana jednostka liniowa) o wzorze

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

Jest ona ciągła, ale nieróżniczkowalna w punkcie 0. Jej pochodna dla $z < 0$ wynosi zero. Jej atutem jest szybkość przetwarzania. Nie ma ona maksymalnej wartości wyjściowej. W zadaniach regresji bywa wykorzystywany „wygładzony” wariant funkcji ReLU, czyli funkcja **softplus**:

$$\text{softplus}(z) = \log(1 + \exp(z))$$

Na rysunku 1.4 przedstawiono popularne funkcje aktywacji wraz z ich pochodnymi.



Rysunek 1.4: Przykładowe funkcje aktywacji wraz z pochodnymi

Źródło: [6]

1.2. Sieci splotowe

Sieci splotowe, nazywane również **splotowymi sieciami neuronowymi** (ang. *convolutional neural networks*, CNN), są rodzajem sieci neuronowych służących do przetwarzania danych o znanej topologii siatki. Przykładem takich danych są szeregi czasowe, które można uznać za jednowymiarową siatkę z próbkami w regularnych odstępach czasu, oraz dane graficzne, które można interpretować jako dwuwymiarową siatkę pikseli. Nazwa sieci splotowych pochodzi od wykorzystywanego przez te sieci działania matematycznego nazywanego **splotem** (konwolucją). Można powiedzieć, że sieci splotowe to po prostu sieci neuronowe, które w przynajmniej jednej z warstw zamiast ogólnego mnożenia macierzy wykorzystują splot [7].

1.3. Czym jest autoenkoder

Autoenkoder (czasem także nazywany autokoderem, z ang. *autoencoder*, *auto-encoder*) jest specjalnym typem sieci neuronowej, która jest przeznaczona głównie do kodowania danych wejściowych do skompresowanej i znaczącej reprezentacji, a następnie dekodowania ich z powrotem w taki sposób, aby zrekonstruowane dane były jak najbardziej podobne do oryginalnych [2]. Autoenkodery uczą się gęstych reprezentacji danych, tzw. reprezentacji ukrytych (ang. *latent representations*) lub kodowań (ang. *codings*) bez jakiegokolwiek formy nadzorowania (tzn. zbiór danych nie zawiera etykiet). Wyjściowe kodowania zazwyczaj mają mniejszą wymiarowość od danych wejściowych, dzięki czemu autoenkodery mogą z powodzeniem służyć do redukcji wymiarowości. Mają też zastosowanie w **modelach generatywnych** (ang. *generative models*), które potrafią losowo generować nowe dane przypominające zbiór uczący. Jeszcze lepszej jakości dane można uzyskać przy użyciu **generatywnych sieci przeciwstawnych**, czyli GAN (ang. *Generative Adversarial Networks*). Sieci GAN są często stosowane w zadaniach zwiększania rozdzielczości obrazu, koloryzowania, rozbudowanego edytowania zdjęć, przekształcania prostych szkiców w realistyczne obrazy, dogenerowywania danych służących do uczenia innych modeli, generowania innych typów danych np. tekstowych, dźwiękowych, itd. [6].

(Chapter 14 Autoencoders [7])

An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer h that describes a code used to represent the input. The network may be viewed as consisting of two parts: an encoder function $h = f(x)$ and a decoder that produces a reconstruction $r = g(h)$. This architecture is presented in figure 14.1. If an autoencoder succeeds in simply learning to set $g(f(x)) = x$ everywhere, then it is not especially useful. Instead, autoencoders are designed to be unable to learn to copy perfectly. Usually they are restricted in ways that allow them to copy only approximately, and to copy

only input that resembles the training data. Because the model is forced to prioritize which aspects of the input should be copied, it often learns useful properties of the data.

Rodzaje autoenkoderów:

- niedopełniony (undercomplete)
- stosowy (stacked) lub inaczej głęboki (deep)
- splotowy (convolutional)
- rekurencyjny (recurrent)
- odszumiający (stacked denoising)
- rzadki (sparse)
- wariacyjny (variational)

Rozdział 2

Przykłady zastosowań autoenkoderów

2.1. Analiza PCA za pomocą autoenkodera niedopełnionego

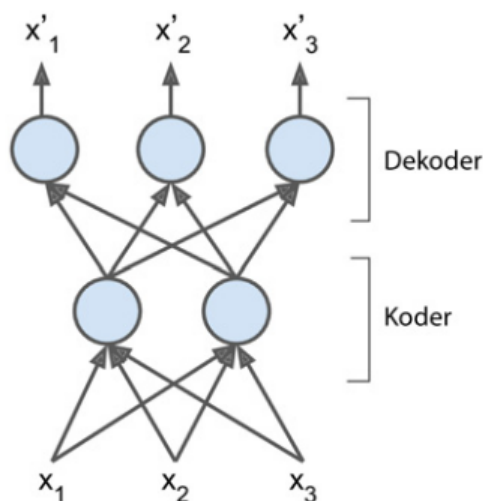
Jeśli autoenkoder korzysta jedynie z liniowych funkcji aktywacji, a funkcją kosztu jest błąd średniokwadratowy (MSE), to będzie on przeprowadzał analizę składowych głównych PCA [6]. Na rysunku 2.1 widoczny jest kod w języku Python służący do utworzenia takiego autoenkodera, który przeprowadzi analizę PCA na zbiorze trójwymiarowym, rzutując go na przestrzeń dwuwymiarową. Można zauważyć podział autoenkodera na dwie części - enkoder i dekodery. Enkoder przyjmuje dane wejściowe o wymiarze 3, a na wyjściu pojawiają się dane dwuwymiarowe. W przypadku dekodera jest odwrotnie - dane na wejściu są dwuwymiarowe, a na wyjściu trójwymiarowe. Obydwie części są modelami sekwencyjnymi z jedną warstwą gęstą, a cały autoenkoder jest modelem sekwencyjnym, w którym po enkoderze występuje dekodery.

```
encoder = keras.models.Sequential([keras.layers.Dense(2, input_shape=[3])])
decoder = keras.models.Sequential([keras.layers.Dense(3, input_shape=[2])])
autoencoder = keras.models.Sequential([encoder, decoder])

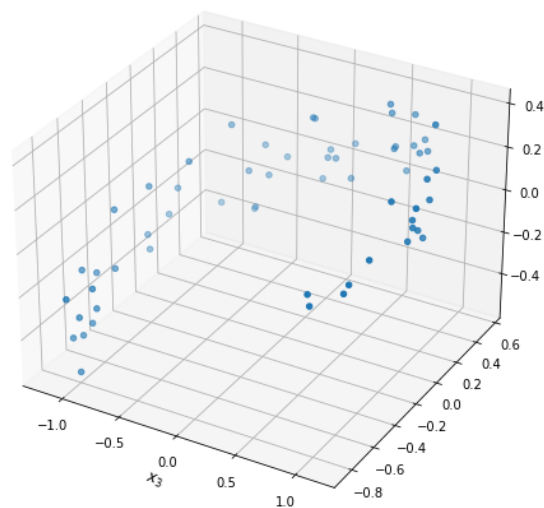
autoencoder.compile(loss="mse", optimizer=keras.optimizers.SGD(learning_rate=1.5))
```

Rysunek 2.1: Tworzenie autoenkodera niedopełnionego przeprowadzającego PCA

Źródło: Opracowanie własne na podstawie [6]

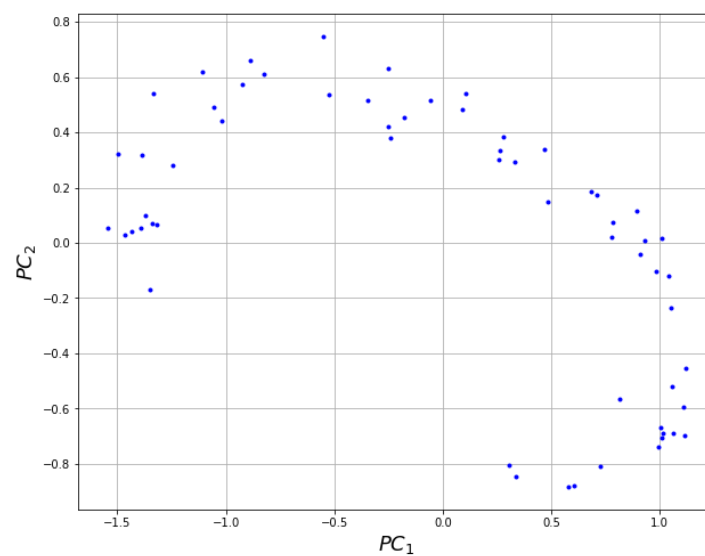
**Rysunek 2.2:** Architektura autoenkodera niedopełnionego*Źródło:* [6]

Opisany autoenkoder zostanie wyuczony na wygenerowanym zbiorze trójwymiarowych danych. Warto zwrócić uwagę, że zbiór uczący stanowi jednocześnie dane wejściowe i dane docelowe. Po wytrenowaniu modelu, używamy enkodera do zakodowania danych wejściowych, czyli do rzutowania ich na przestrzeń dwuwymiarową. Na rysunku 2.3 zaprezentowany jest trójwymiarowy zbiór danych. Rysunek 2.4 przedstawia wynik działania autoenkodera, a więc dane rzutowane na przestrzeń 2D. Z kolei na rysunku 2.5 widoczne jest rzutowanie uzyskane przy użyciu „zwykłego” algorytmu PCA. Na podstawie tych wykresów można stwierdzić, że rzutowanie uzyskane przy pomocy autoenkodera oraz PCA jest takie samo, jedynie układ współrzędnych jest odwrócony o 180 stopni.



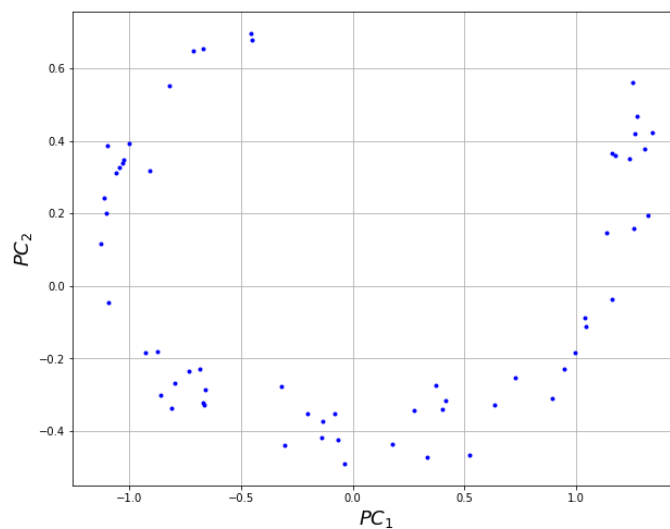
Rysunek 2.3: Wygenerowane dane trójwymiarowe

Źródło: Opracowanie własne na podstawie [6]

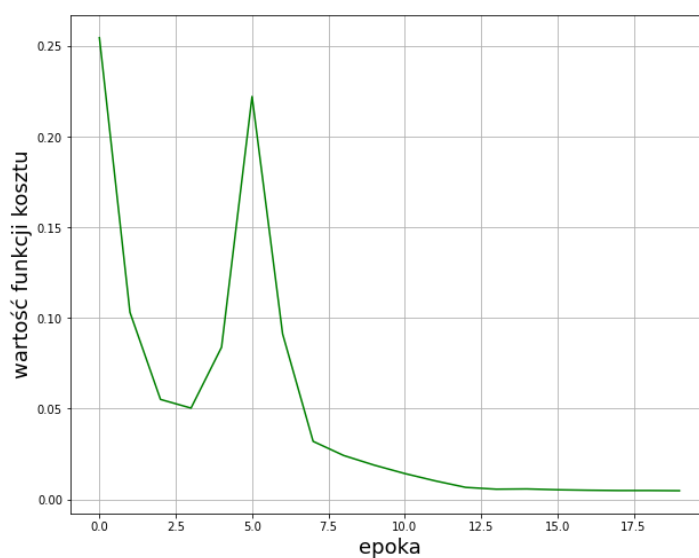


Rysunek 2.4: Rzutowanie dwuwymiarowe przy użyciu autoenkodera, zachowujące maksymalną wariancję

Źródło: Opracowanie własne na podstawie [6]



Rysunek 2.5: Rzutowanie dwuwymiarowe przy użyciu PCA, zachowujące maksymalną wariancję
Źródło: Opracowanie własne na podstawie [6]



Rysunek 2.6: Wartość funkcji straty w kolejnych epokach podczas uczenia modelu
Źródło: Opracowanie własne

2.2. Kolejny przykład

Podsumowanie i wnioski

Bibliografia

- [1] Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*, Springer International Publishing
- [2] Bank D., Koenigstein N., Giryas R. (2021), *Autoencoders*, arXiv:2003.05991v2
- [3] Chollet F., Allaire J. J. (2019) *Deep Learning. Praca z językiem R i biblioteką Keras*, Helion SA
- [4] Edureka, Autoencoders Tutorial. Autoencoders in Deep Learning. Tensorflow Training https://www.youtube.com/watch?v=nTt_ajul8NY
- [5] Ertel W. (2017) *Introduction to Artificial Intelligence. Second Edition*, Springer International Publishing
- [6] Géron A. (2020) *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Wydanie II*, Helion SA
- [7] Goodfellow I., Bengio Y., Courville A. (2018), *Deep Learning. Systemy uczące się*, PWN, Warszawa
- [8] Osinga D. (2019) *Deep Learning. Receptury*, Helion SA
- [9] Skansi S. (2018) *Introduction to Deep Learning. From Logical Calculus to Artificial Intelligence*, Springer International Publishing

Spis rysunków

1.1	Przykładowe sztuczne sieci neuronowe rozwiązujące proste zadania logiczne	7
1.2	Struktura sztucznego neuronu, który stosuje funkcję skokową f na ważonej sumie sygnałów wejściowych	7
1.3	Perceptron z trzema neuronami wejściowymi i trzema wyjściami	8
1.4	Przykładowe funkcje aktywacji wraz z pochodnymi	10
2.1	Tworzenie autoenkodera niedopełnionego przeprowadzającego PCA	13
2.2	Architektura autoenkodera niedopełnionego	14
2.3	Wygenerowane dane trójwymiarowe	15
2.4	Rzutowanie dwuwymiarowe przy użyciu autoenkodera, zachowujące maksymalną wariancję	15
2.5	Rzutowanie dwuwymiarowe przy użyciu PCA, zachowujące maksymalną wariancję . . .	16
2.6	Wartość funkcji straty w kolejnych epokach podczas uczenia modelu	16

Spis tabel

Załączniki

1. Płyta CD z niniejszą pracą w wersji elektronicznej.

Streszczenie (Summary)

Przegląd autoenkoderów stosowanych w nienadzorowanym uczeniu maszynowym

An overview of autoencoders used in unsupervised machine learning