

Deep Learning na przykładzie Deep Belief Networks

Jan Karwowski

Zakład Sztucznej Inteligencji i Metod Obliczeniowych
Wydział Matematyki i Nauk Informatycznych PW

20 V 2014



- Głębokość – maksymalna długość ścieżki od wejścia do wyjścia sieci (liczba neuronów)
- Początkowo ('50, '60) zaniechana ze względu na brak możliwości obliczeniowych
- Wzrost popularności w ostatnich 10 latach
- Trudność w uczeniu głębokich układów klasycznymi algorytmami
- <http://deeplearning.net>



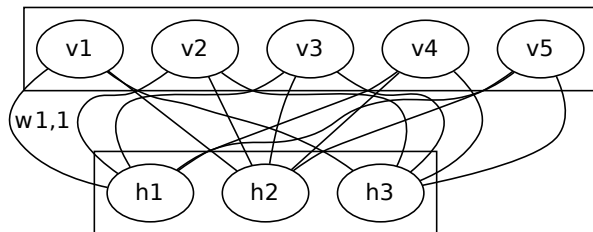
- Rzeczywiste układy biologiczne
 - Ośrodek widzenia w mózgu kota ma około 10 warstw
- Istnieją funkcje trudne do reprezentacji w płaskim modelu (parzystość)
- Sieci wielowarstwowe mogą łączyć preprocessing i klasyfikację/regresję



Restricted Boltzman Machine

- Neurony binarne
- Prawdopodobieństwo pobudzenia neuronu
 $P(h_k = 1|v) = \sigma(b_k + \sum_i w_{i,k}v_i)$
- Każde połączenie działa w obie strony
- Energia układu

$$E(v, h) = - \sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i h_j w_{ij}$$



Contrastive Divergence

- Geoffrey E Hinton. “Training products of experts by minimizing contrastive divergence”. In: *Neural computation* 14.8 (2002), pp. 1771–1800
- Geoffrey Hinton. *A practical guide to training restricted Boltzmann machines*. 2010
- Tijmen Tieleman. “Training restricted Boltzmann machines using approximations to the likelihood gradient”. In: *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, pp. 1064–1071



Contrastive Divergence

v_1, \dots, v_n – warstwa widoczna, h_1, \dots, h_m – warstwa ukryta, $W \in \mathbb{M}_n^m$ – macierz wag połączeń, η – wsp. uczenia

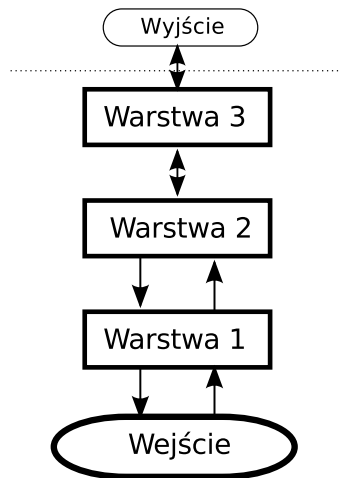
Algorithm 1 Pojedyncza epoka dla pojedynczego wzorca

```
1: function CONTRASTIVEDIVERGENCE( $V, W$ )
2:    $v \leftarrow V$ 
3:    $H \leftarrow h \leftarrow \text{SAMPLEHIDDEN}(v, W)$ 
4:   for  $i \leftarrow 1..n$  do
5:      $v \leftarrow \text{SAMPLEVISIBLE}(h, W)$ 
6:      $h \leftarrow \text{SAMPLEHIDDEN}(v, W)$ 
7:   end for  $\Delta W \leftarrow \eta( \text{OUTERPRODUCT}(V, H) - \text{OUTERPRODUCT}(v, h) )$ 
8:   return  $W + \Delta W$ 
9: end function
```

- Pomijamy wpływ zmiany wag na poprzednie próbkowania



Deep Belief Network

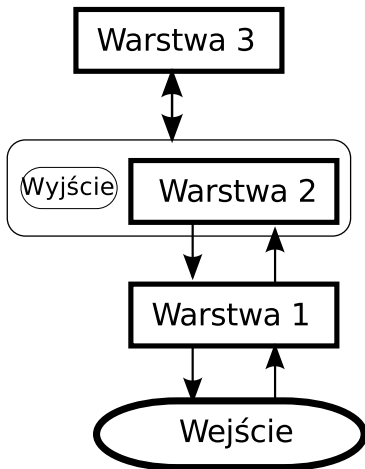


- Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. “A fast learning algorithm for deep belief nets”. In: *Neural computation* 18.7 (2006), pp. 1527–1554
- model generatywny



- 1: Uczenie każdej warstwy połączeń osobno za pomocą CD, w niższej warstwie wartości rzeczywiste – prawdopodobieństwa wystąpienia 1, w wyższej próbkowane 0,1
- 2: Uczenie części klasyfikacyjnej jako elementu rozszerzającego przedostatnią warstwę neuronów
- 3: **for** $i \leftarrow 1..$ liczba epok **do**
- 4: Próbkowanie w górę
- 5: Naprzemienne próbkowanie w najwyższej warstwie
- 6: Propagacja w dół
- 7: Gradientowa poprawa wszystkich wag
- 8: **end for**





Warstwę 2 i wyjście traktujemy jakby była to jedna warstwa.



Wariant 1:

- 1 Z wejścia próbkuj wartości wyższych warstw, aż do warstwy z dwukierunkowymi połączeniami
- 2 W obrębie górnej RBM wykonuj naprzemienne próbkowanie, aż do osiągnięcia stanu równowagi
- 3 Zainicjuj wartości wszystkich etykiet klas wyjścia na 0.1 i wykonuj naprzemienne próbkowanie w warstwie rozpoznającej aż do osiągnięcia równowagi

Wariant 2:

- 1 Z wejścia próbkuj wartości wyższych warstw, aż do warstwy z dwukierunkowymi połączeniami
- 2 Dla każdej klasy wykonaj: ustaw etykietę danej klasy na 1, pozostałe na 0, wykonuj naprzemienne próbkowanie warstwy etykiet z dolną warstwą górnej RBM i górnej warstwy RBM, aż do osiągnięcia równowagi
- 3 Wybierz klasę, dla której końcowa energia układu była najmniejsza



Zbiór MNIST

Version of MNIST task	Learning algorithm	Test error %
permutation-invariant	Our generative model 784 \rightarrow 500 \rightarrow 500 \leftrightarrow 2000 \leftrightarrow 10	1.25
permutation-invariant	Support Vector Machine degree 9 polynomial kernel	1.4
permutation-invariant	Backprop 784 \rightarrow 500 \rightarrow 300 \rightarrow 10 cross-entropy & weight-decay	1.51
permutation-invariant	Backprop 784 \rightarrow 800 \rightarrow 10 cross-entropy & early stopping	1.53
permutation-invariant	Backprop 784 \rightarrow 500 \rightarrow 150 \rightarrow 10 squared error & on-line updates	2.95
permutation-invariant	Nearest Neighbor All 60,000 examples & L3 norm	2.8
permutation-invariant	Nearest Neighbor All 60,000 examples & L2 norm	3.1
permutation-invariant	Nearest Neighbor 20,000 examples & L3 norm	4.0
permutation-invariant	Nearest Neighbor 20,000 examples & L2 norm	4.4

Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. “A fast learning algorithm for deep belief nets”. In: *Neural computation* 18.7 (2006), pp. 1527–1554



Generatywna część modelu



Figure 9: Each row shows 10 samples from the generative model with a particular label clamped on. The top-level associative memory is initialized by an up-pass from a random binary image in which each pixel is on with a probability of 0.5. The first column shows the results of a down-pass from this initial high-level state. Subsequent columns are produced by 20 iterations of alternating Gibbs sampling in the associative memory.

- Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets". In: *Neural computation* 18.7 (2006), pp. 1527–1554
- Próbkowanie w najwyższej warstwie, propagacja jak w perceptronie w niższych.



- Deep Boltzman Machine
- Autoencoder
- Convolutional Neural Network
- Deep SVM



+

- Dobre działanie dla problemów z rozpoznawaniem obrazów, dla dużej liczby danych wejściowych
- Wykorzystanie znanych wcześniej płytkich modeli

-

- Duża złożoność, w szczególności pamięciowa
- Trudna analiza teoretyczna

