Foundations of Machine Learning
Neural Networks

Kate Farrahi

ECS Southampton

November 19, 2020

References

- Pattern Recognition and Machine Learning by Christopher Bishop

References

- Pattern Recognition and Machine Learning by Christopher Bishop
- Michael Nielson's online book http://neuralnetworksanddeeplearning.com

References

- Pattern Recognition and Machine Learning by Christopher Bishop
- Michael Nielson's online book
  `http://neuralnetworksanddeeplearning.com`
- Deep Learning by Ian Goodfellow, Y. Bengio, and A. Courville
  `http://www.deeplearningbook.org`

References

- Pattern Recognition and Machine Learning by Christopher Bishop
- Michael Nielson's online book
  http://neuralnetworksanddeeplearning.com

- Deep Learning by Ian Goodfellow, Y. Bengio, and A. Courville
  http://www.deeplearningbook.org
- Step by Step Example of Backpropagation:
  http://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example

References

- Pattern Recognition and Machine Learning by Christopher Bishop
- Michael Nielson's online book
  http://neuralnetworksanddeeplearning.com

- Deep Learning by Ian Goodfellow, Y. Bengio, and A. Courville
  http://www.deeplearningbook.org
- Step by Step Example of Backpropagation:
  http://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example

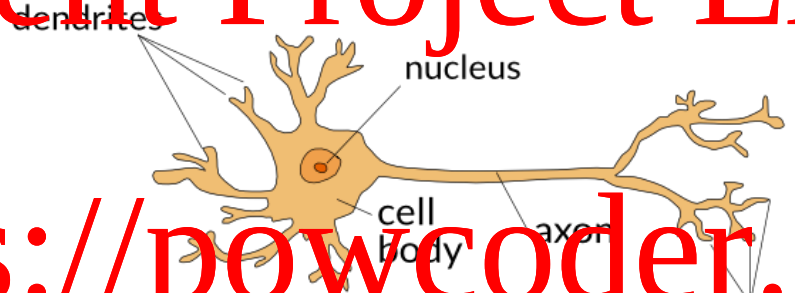
The Neuron

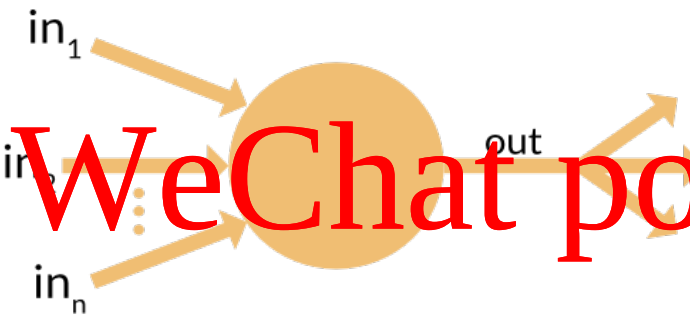
The Human Brain

- Highly complex, non-linear, and parallel "computer"
- Structural constituents: neurons
- The structure of the brain is extremely complex and not fully understood
- Billions of nerve cells (neurons) and trillions of interconnections in the human brain
- Scientists tried to mimic the brain's behaviour in proposing the artificial neural network (ANN)
- The human brain is the inspiration for ANNs though we cannot say ANNs actually replicate the brain's behaviour very well, they are extremely simplified
- Great video about the brain
  `https://www.youtube.com/watch?v=nvXuq9jRWKE`

The Neuron

biological neuron





artificial neuron

History of Neural Networks: McCulloch-Pitts Model

- 1943 McCulloch and Pitts introduced the first model of an extremely simple artificial neuron.
- The inputs and outputs could be either a zero or a one.
- They introduced the idea of an excitatory and inhibitory potential using weights (+/-).

- Each input is weighed and the summed activation is either transmitted (output of 1) or not (output of 0).
- The McCulloch-Pitts model lacked a mechanism for learning, which was crucial for it to be usable for AI.

- Link to the Original Paper `https://link.springer.com/article/10.1007%2FBF02478259`

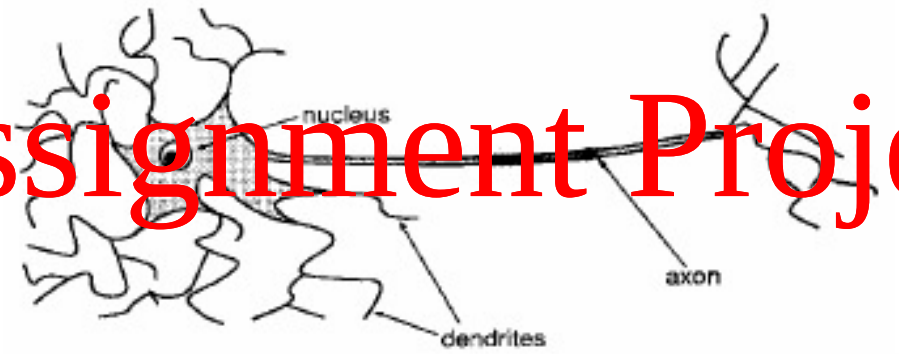
History of the Perceptron

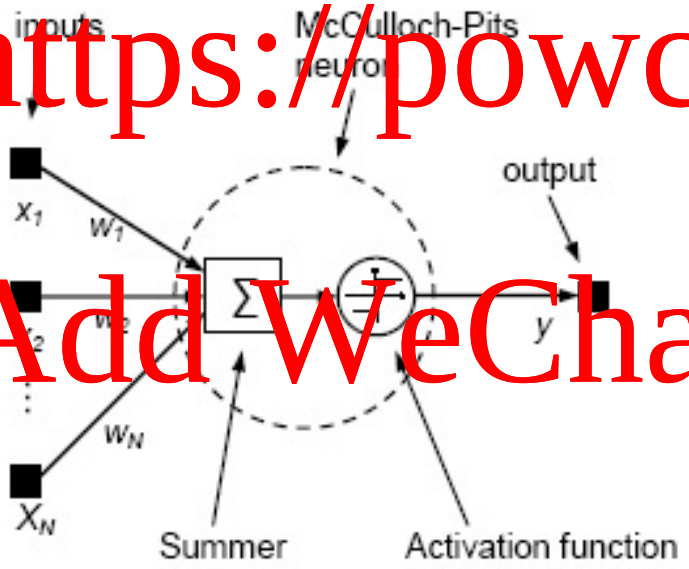
- 1957 Rosenblatt introduced the perceptron which was an electronic device constructed using biological principles and showed the ability to learn.
- 1962 Rosenblatt wrote a book about the Perceptron and received international recognition.
- 1969 Marvin Minsky and Seymour Papert published the book "Perceptrons" which proved some limitations of the perceptron (that linear functions cannot model non-linears ones) having a big effect on the community.

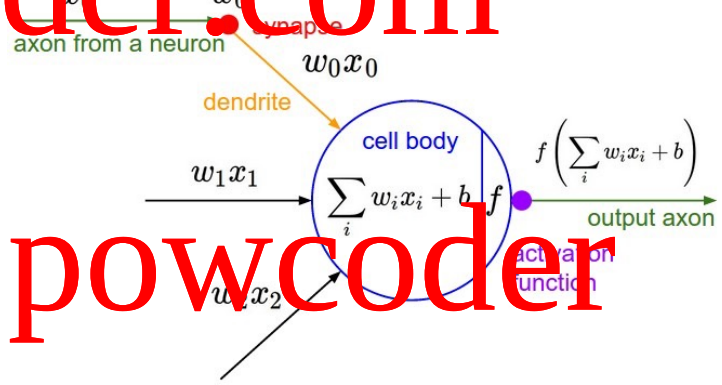
History of the Perceptron

- Initially the perceptron seemed promising, but it was quickly shown that perceptrons could not be used to classify many classes of patterns.
- This caused the field of neural networks to stagnate for many years before it was recognised that a feedforward neural network with two or more layers (multilayer perceptrons) had far greater power.
- The popularity of neural networks resurged in the 1980s.
- Today deep learning is state of the art for many applications in machine learning.

The Perceptron







Source: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

The Perceptron Algorithm

Begin Initialize

Set all of the weights w_i to small random numbers

Training

For T iterations (or until the convergence criteria is met):

For each input vector x_j:

Compute the activation of each neuron i:

$$y_j = f(\sum_{i=0}^{m} w_i x_{ij}) = \begin{cases} 1 & \text{if } \sum_{i=0}^{m} w_i x_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Update each weight as follows:

$$w_i \leftarrow w_i - \eta(y_j - t_j) \cdot x_{ij} \quad (2)$$

Example

- Solve the logical AND function using the perceptron algorithm.

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?
  - What are some suitable convergence criteria?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?
  - What are some suitable convergence criteria?
  - Is there more than one possible solution?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?
  - What are some suitable convergence criteria?
  - Is there more than one possible solution?
  - What happens if you set η to a very large number?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?
  - What are some suitable convergence criteria?
  - Is there more than one possible solution?
  - What happens if you set η to a very large number?
  - What happens if you set η to a very small number?

Example

- Solve the logical AND function using the perceptron algorithm
  - Given $b = 1, w_1 = 0, w_2 = 0, \eta = 0.1$, find a solution
  - After how many iterations did the perceptron converge?
  - Do you need a bias term?
  - What are some suitable convergence criteria?
  - Is there more than one possible solution?
  - What happens if you set η to a very large number?
  - What happens if you set η to a very small number?

Example

Solve the logical XOR function using the perceptron algorithm.