Neural Networks

Shyue Ping Ong

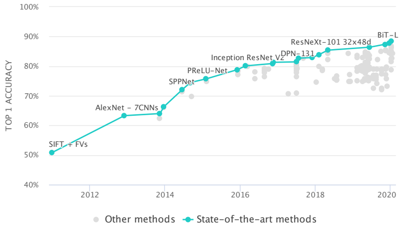
University of California, San Diego

NANO281

Overview

Preliminaries

- Neural networks/deep learning has gotten a lot of hype in recent years.
- In many areas, they have outperformed many traditional ML methodologies.

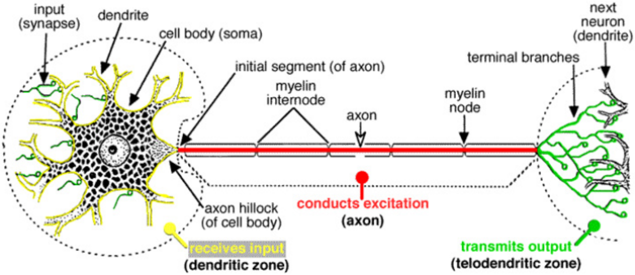


Artificial Neural Network

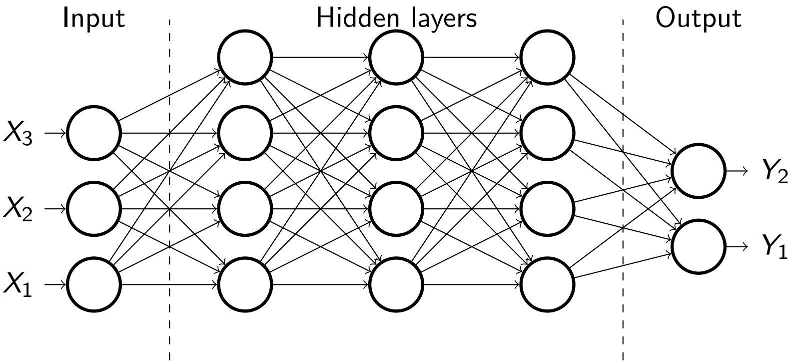
- An artificial neural network (NN) is a learning algorithm that is (very) loosely based on the structure of the brain.

Universal Approximation Theorem[1]

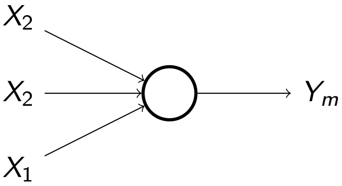
A feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions under mild assumptions on the activation function.



Neural Networks



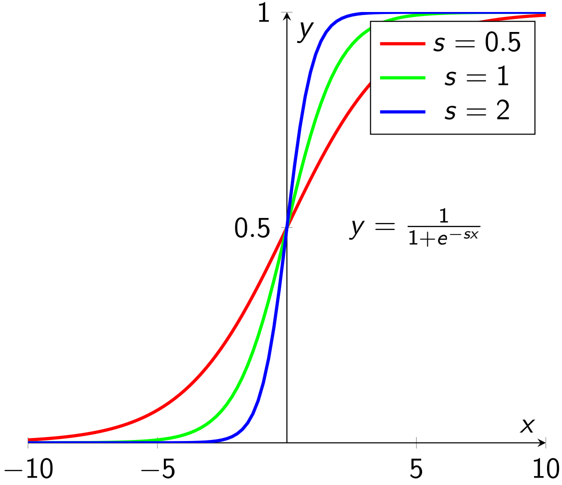
Neuron



$$Y_m = \sigma(\alpha_{0m} + \alpha_m^T \mathbf{X})$$

- Output of each neuron is a linear function of the inputs.
- In hidden layers, the output is passed through an *activation function* σ.

Sigmoid Activation Function



Fitting a Neural Network

- Model *weights* (α in the linear functions) are fitted by *back-propogation*, basically a form of gradient descent.
- Loss functions: squared error for regression, squared error or cross entropy for classification.
- To avoid overfitting, regularization (similar to ridge regression) is typically applied. E.g., *weight decay*:

$$J = \sum \alpha^2$$

Decisions for Neural Networks

- Number of hidden units and layers: generally error on the side of having too many hidden units than too few - flexibility is needed to capture non-linearities in the data.
- Extra weights can be shrunk to zero with appropriate regularization.
- Learning rate is a key parameter in NN as well as other model fitting. The learning rate controls the rate of gradient descent.

Bibliography

Balázs Csanád Csáji.
Approximation with Artificial Neural Networks.
PhD thesis.

The End