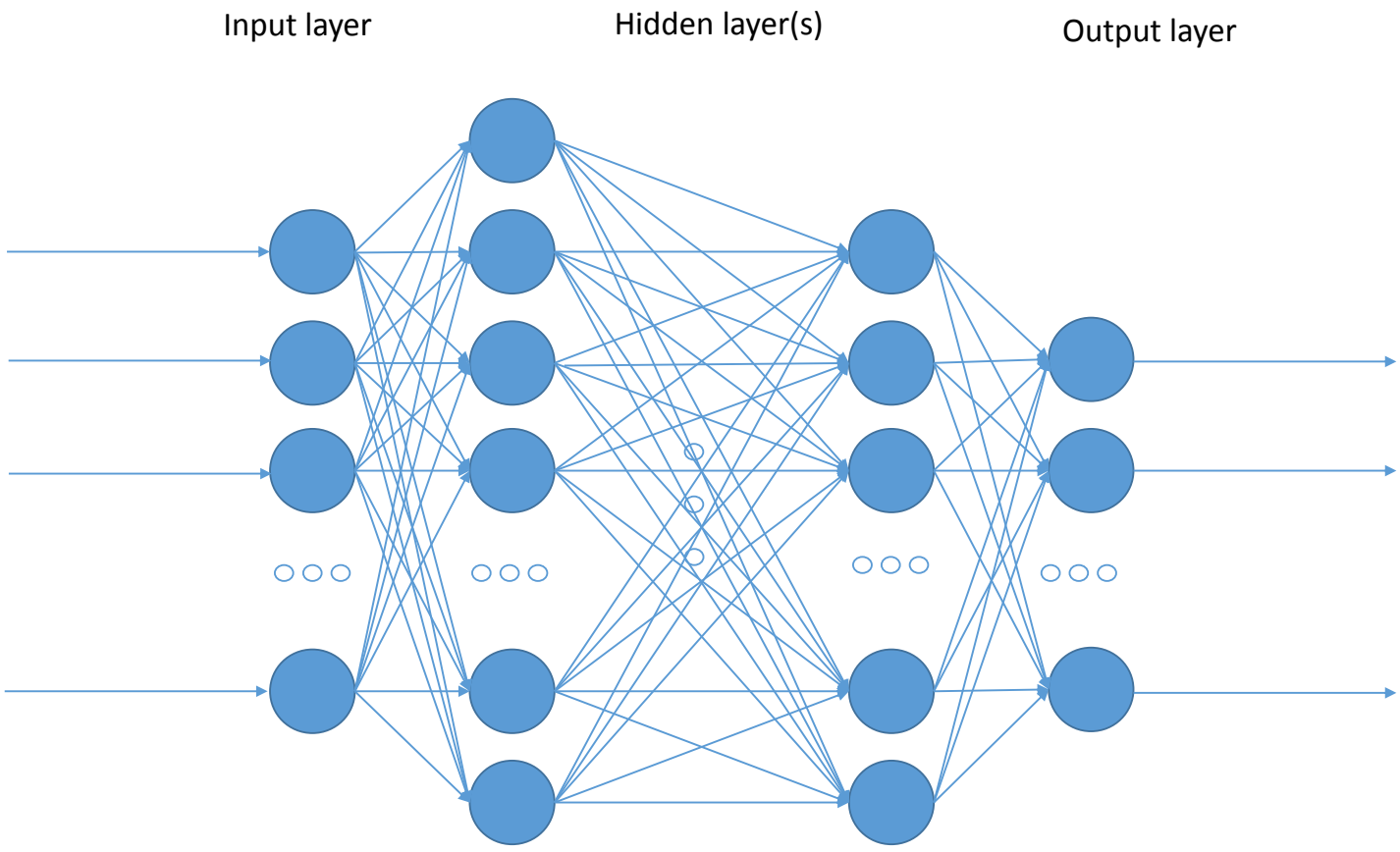
Machine Learning, Lecture 8: Artificial Neural Networks

S. Nõmm

[1]Department of Computer Science, Tallinn University of Technology

26.03.2015

What is *Artificial Neural Network ?*



Biology inspired mathematical abstraction

Artificial Neural Network

- Brain (human or animal) consists of neurons.
- Basic idea is to combine together a number of neurons.
- Perceptron is the mathematical model of a single neuron.
- Perceptron has a number of drawbacks which motivate creation of artificial neural networks
  - Does not provide probabilistic outputs.
  - Does not suited for more than two classes.
  - Learns linear decision boundaries only.

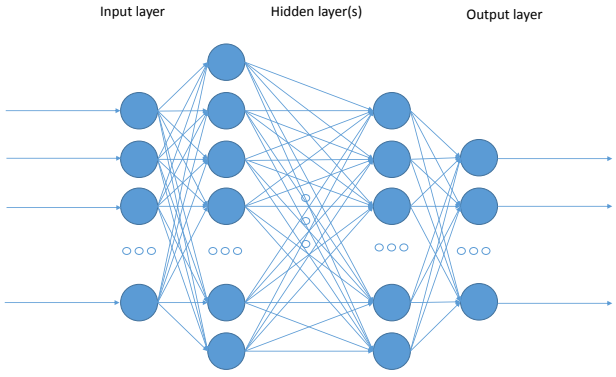
The model of a single neuron

$$f(a) = \text{activation function}; \quad \text{where} \quad a = \sum_{j=1}^{d} w_j x_j = \boldsymbol{w}^T \boldsymbol{x};$$

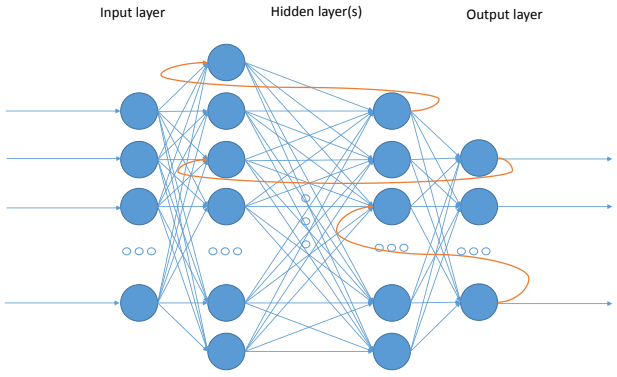
where $\boldsymbol{x} \in \mathbb{R}^d$ is the input vector. The neuron has d weights the same number as inputs.
Note, *activation or saturation* function - sigmoid function (logsig, tansig etc.) Note, some times bias or intercept is added.

$$a = \sum_{j=1}^{d} w_j x_j + b = \boldsymbol{w}^T \boldsymbol{x} + b;$$

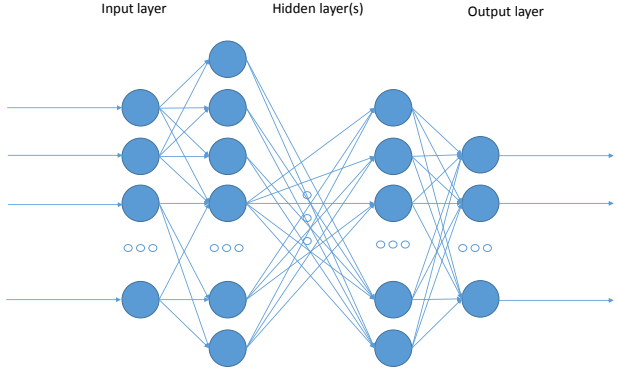
Assembling neurons together



Fully connected feed forward neural network



Recurrent neural network



Restricted connectivity feed forward neural network

ANN

- Modeling of nonlinear processes
- Classification (Next lecture)
- ... .

More practical approach

- Nonlinear process and function approximation
- Pattern recognition and classification
- Clustering
- Time Series and Dynamic Systems

Modeling of nonlinear processes

- Topology (network connectivity)
- Number of the neurons on each level
- Activation function
- training method

ANN

- How many layers do one need?
- How many neurons are necessary?
- Universal approximation theorem:
  - G.Cybenko (1989) Single hidden layer with finite number of neurons (multilayer perceptron) can approximate continuous functions on a compact subsets of $\mathbb{R}^n$. Some weak assumptions about activation function were made. Algorithmic aspects were not touched.
  - K.Hornik (1991) Demonstrated importance of the choice of architecture over the choice of activation function.
- While there is no constructive analytic approach to select structure of the network there exist number of rules which may be used.

Most popular activation functions

Activation function some times referred as transfer function.

- logsig - Log - sigmoid transfer function
- tansig - Tan-sigmoid
- purelin - Linear function

Training techniques

- Levenberg-Marquardt backpropagation.
- Bayesian regularization backpropagation.
- Scaled conjugate gradient backpropagation.
- Resilient backpropagation.

Finishing touches

- Initialization
- Stopping criteria
- Training time is measured in *epochs*. An epoch is a measure of the number of times all of the training vectors are used once to update the weights.
- What does NN-based nonlinear function looks like? Relatively simple case

$$y = \left(\sum_{k=1}^{l} c_k \phi_k \left(\sum_{j=1}^{m} \omega_{k,j}^T x \right) \right)$$

l is the number of neurons on hidden layer, m is the number of inputs, ϕ_i are activation functions.

Example ?

Let us continue in the computer class