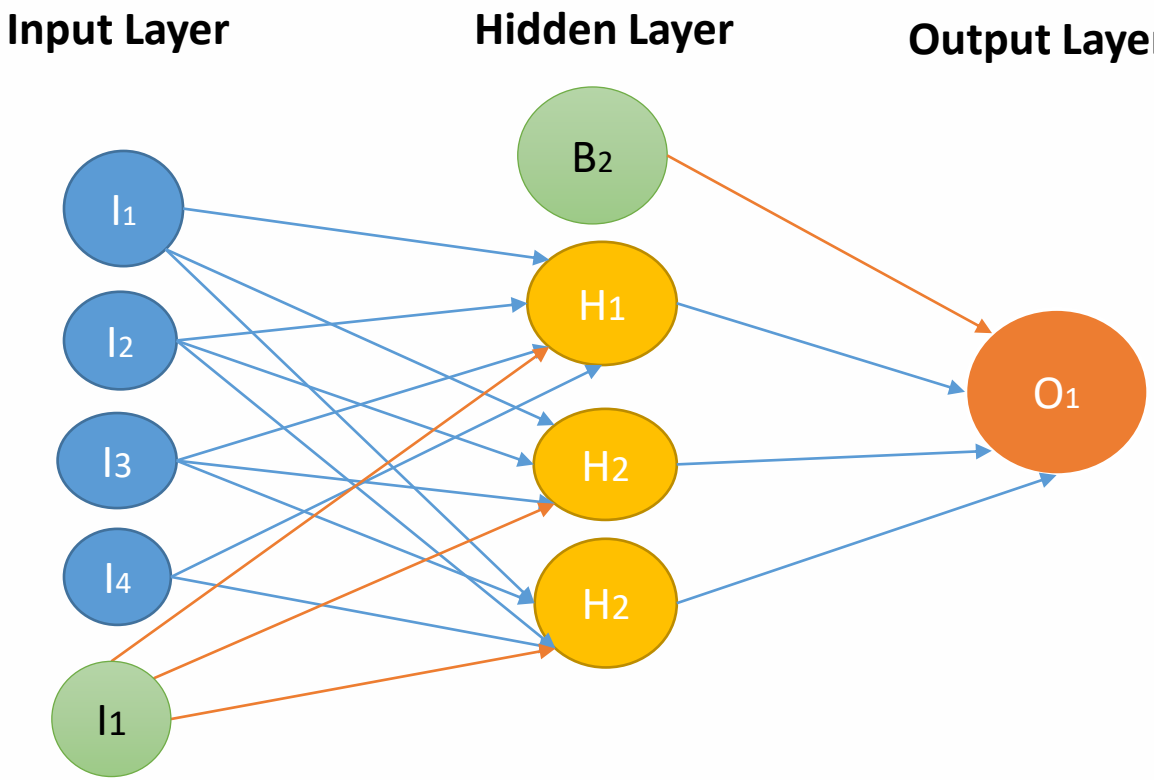
Sane's STATS Academy of Statistics

Neural Networks

Feed Forward

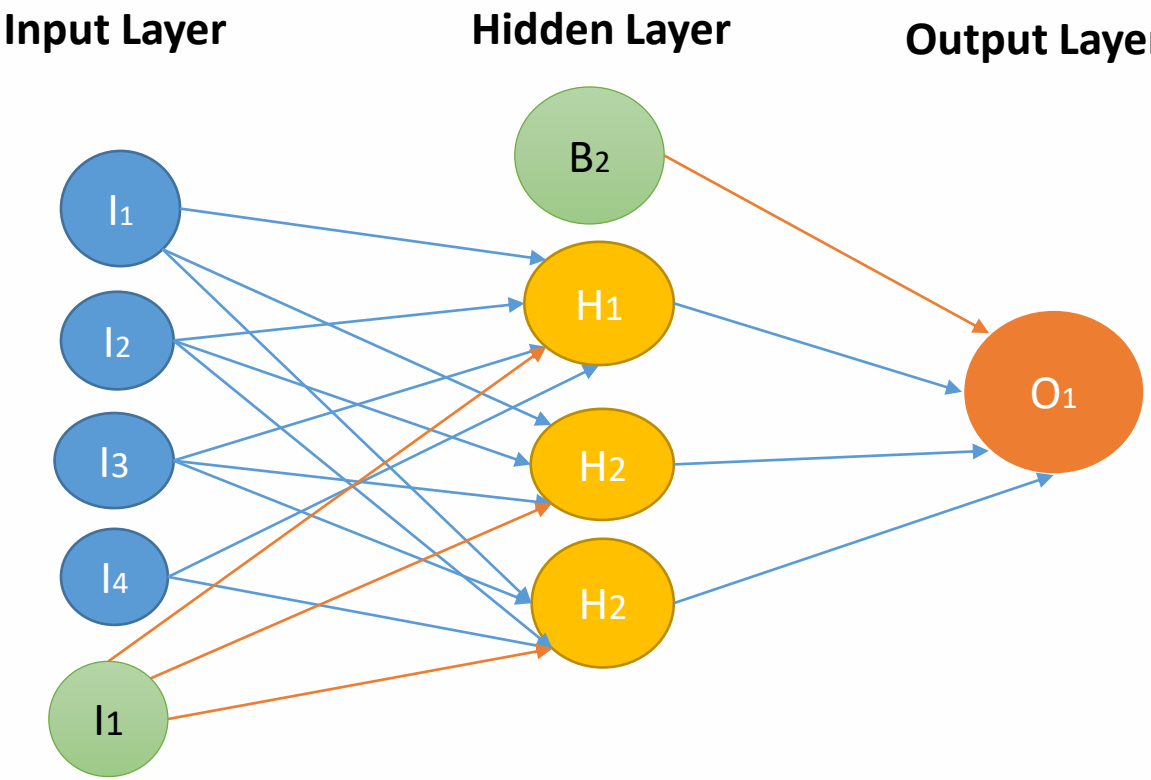
Neural Network

- A neural network is constructed from a number of interconnected nodes called neurons.
- Neurons are arranged into input layer, hidden layer and output layer.
- Input layer corresponds to our predictors / features and Output layer to our response variable(s).



Multi-Layer Perceptron

- The neural network with input layer, one or more hidden layers and one output layer is called multi-layer perceptron (MLP).
- MLP given below has 4 input nodes, 3 hidden nodes and one output node



Functioning of MLP

- Input layer neurons receive incoming information from the data which they process and distribute it to the hidden layers.
- That information, in turn is processed by hidden layers and is passed to the output neurons.
- The information in this artificial neural network is processed in terms of one activation function. This function actually imitates the brain neurons.
- Each neuron contains a value of activation functions and a threshold value.
- The threshold value is the minimum value that must be possessed by the input so that it can be activated.
- The task of the neuron is perform a weighted sum of all the input signals and apply activation function on the sum before passing it to the next(hidden or output) layer.

Weighted Sum

- Say that, we have values a_1, a_2, a_3, a_4 for input and weights as w_1, w_2, w_3, w_4 as the input to one of the hidden layer neuron say n_j, then the weighted sum is represented as

$$S_j = \sum_{i=1}^{4} w_i a_i + b_j$$

where b_j: bias due to node

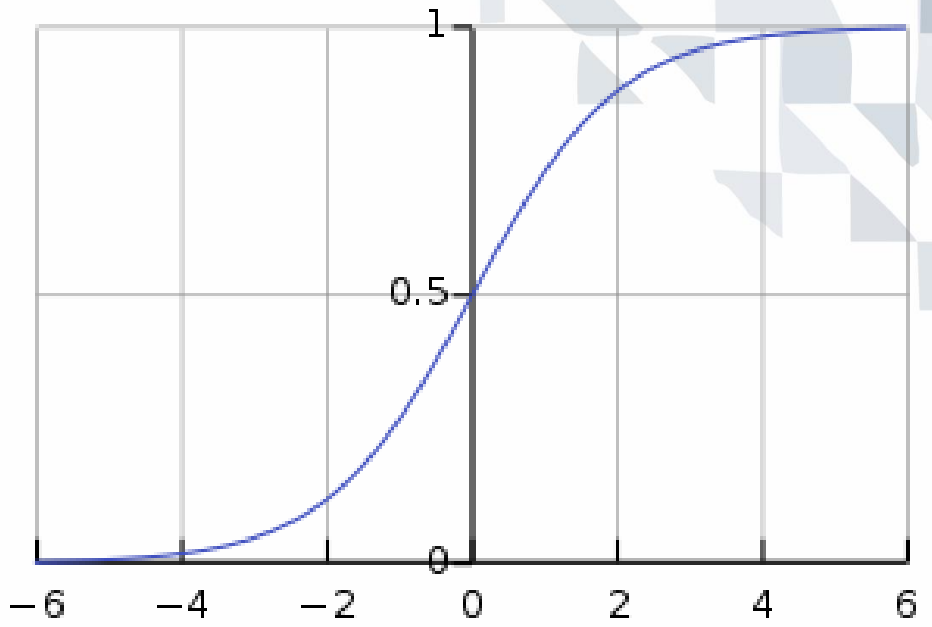
Activation Functions

- These functions are needed to introduce a non-linearity into the network
- Activation function is applied and that output is passed to the next layer
- Possible Functions
  - Sigmoid
  - Hyperbolic Tangent
  - ReLU
  - Softmax

Sigmoid Function

- It is differentiable
- It produces output between 0 and 1
- Function form is

$$S(x) = \frac{1}{1 + e^{-x}}$$

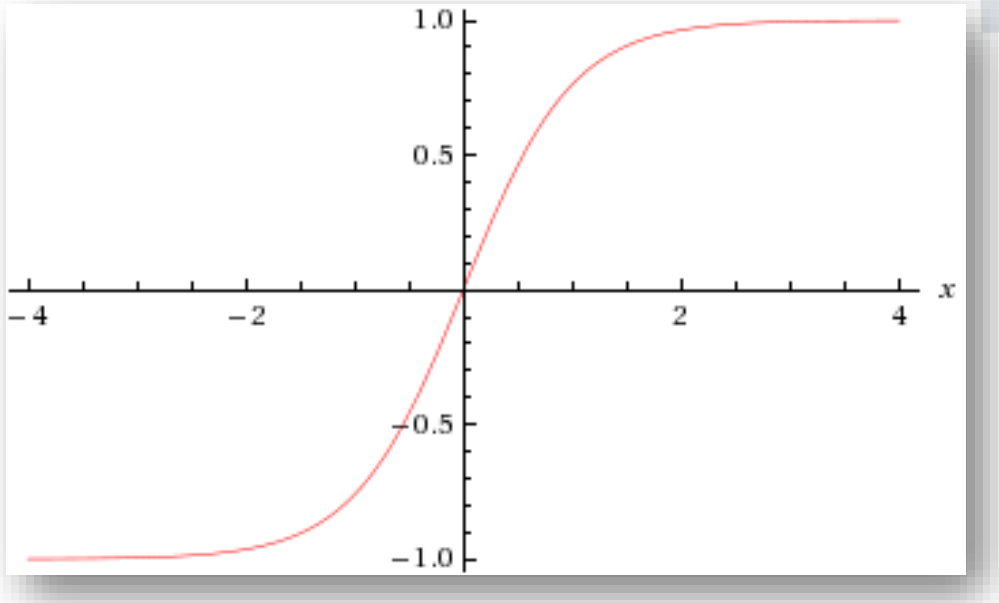


Source: Wikipedia

Hyperbolic Tangent Function

- This is also differentiable
- Produces output between -1 and 1
- Function Form is

$$tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

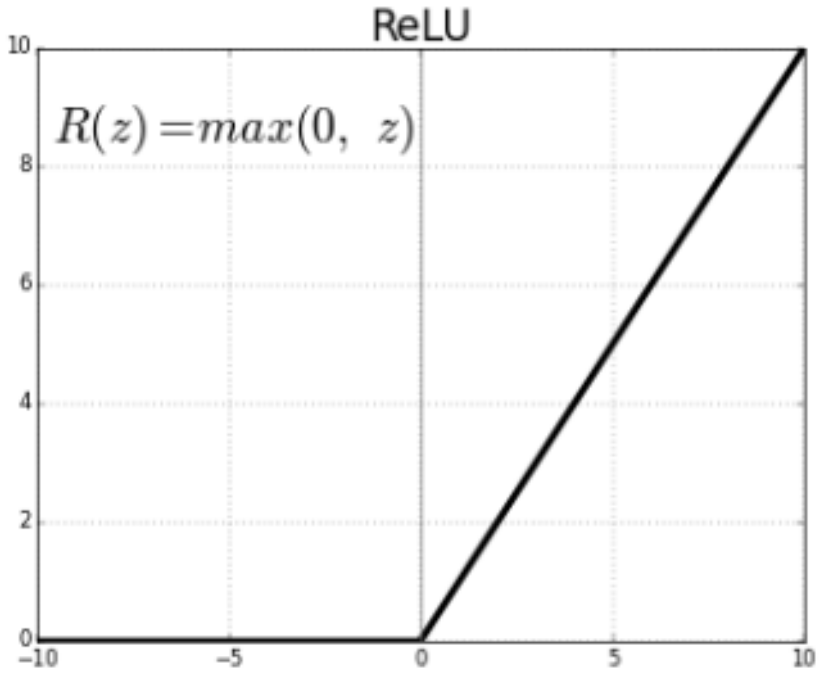


Sane's STATS Academy of Statistics

Rectified Linear Unit(ReLU) Function

- Most popular
- Used widely in deep learning

f(x) = max(0,x)



Softmax Function

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{m} e^{x_j}}$$; m being the number of different categories

- Used for multi-class classification problems
- It is a generalization of sigmoid function
- It produces output between 0 and 1

Learning of Neural Network

- All the learning algorithms of neural network train the network by iteratively modifying the connection weights until the error desired output and the calculated output does not drop to a desired level
- Popularly known algorithm we are going to learn is backpropagation algorithm which uses gradient descent method

Mathematical Significance of Backpropagation

- According to Calculus, chain of functions can be differentiated using the following identity, called the chain rule : f(g(x)) = f'(g(x)) * g'(x).
- If we apply the same chain rule to the computation of the gradient values of the neural network, then it is called backpropagation (reverse-mode differentiation)
- Backpropagation starts with the error in the last layer and traverses back up to the first layer

Neural Network Essentials

- Pass: One pass = one forward pass + one backward pass (we do not count the forward pass and backward pass as two different passes)
- Iterations: Number of passes, each pass using specific batch size number of examples.
- Epoch: one forward pass and one backward pass of all the training observations

Backpropagation Algorithm

1. **Initialization of weights**: Weights are randomly assigned usually between -1 and 1
2. **Feed Forward**: Information(calculated data) is passed on to the next layer starting from input layers to hidden layers to output layers with the help of activation function
3. **Error Assessment**: The output of the network is assessed relative to the existing output. If error is below a permissible value then algorithm is terminated, otherwise proceed to step 4

Backpropagation Algorithm

4. **Propagation**: The error of the output layer is used to modify the weights. The errors are propagated backwards through the network thereby calculating the gradient of change in error
5. **Adjust**: Adjustments are made to the weights using the gradients of change with the goal of reducing the error. The weights and biases are adjusted by a factor based on the derivative of the activation function and the errors. Now Step 2 is executed.

Types of Backpropagation

- Classical Backpropagation
- Momentum Backpropagation
- Resilient Propagation
- ADAM
- And many more

Methods of training network

- Online training: Updates the weights based on gradients calculated from a single training set observation
- Batch training: Updates the weights based on the sum of the gradients over all training set observations
- Batch Size training: Updates the weights based on the sum of some large batch size of training set observations
- Mini-batch training: Same as Batch Size training, only difference is that batch size is small like 32 to 64 only