Part 2

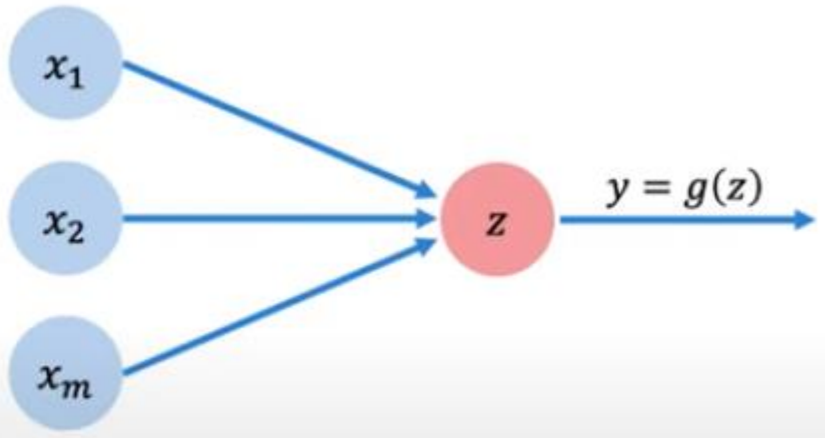
Artificial Neural Networks (ANN)

ESP 3201

Dr Ng Gee Wah

National University of Singapore

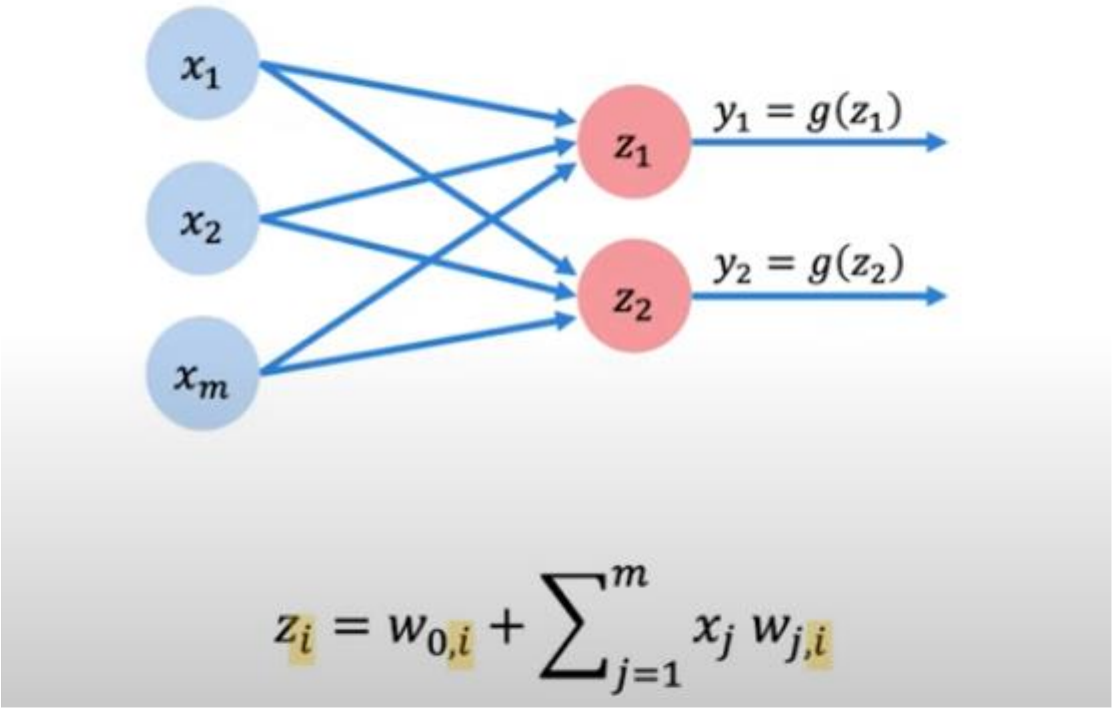
Revisit: The Perceptron: Simplified



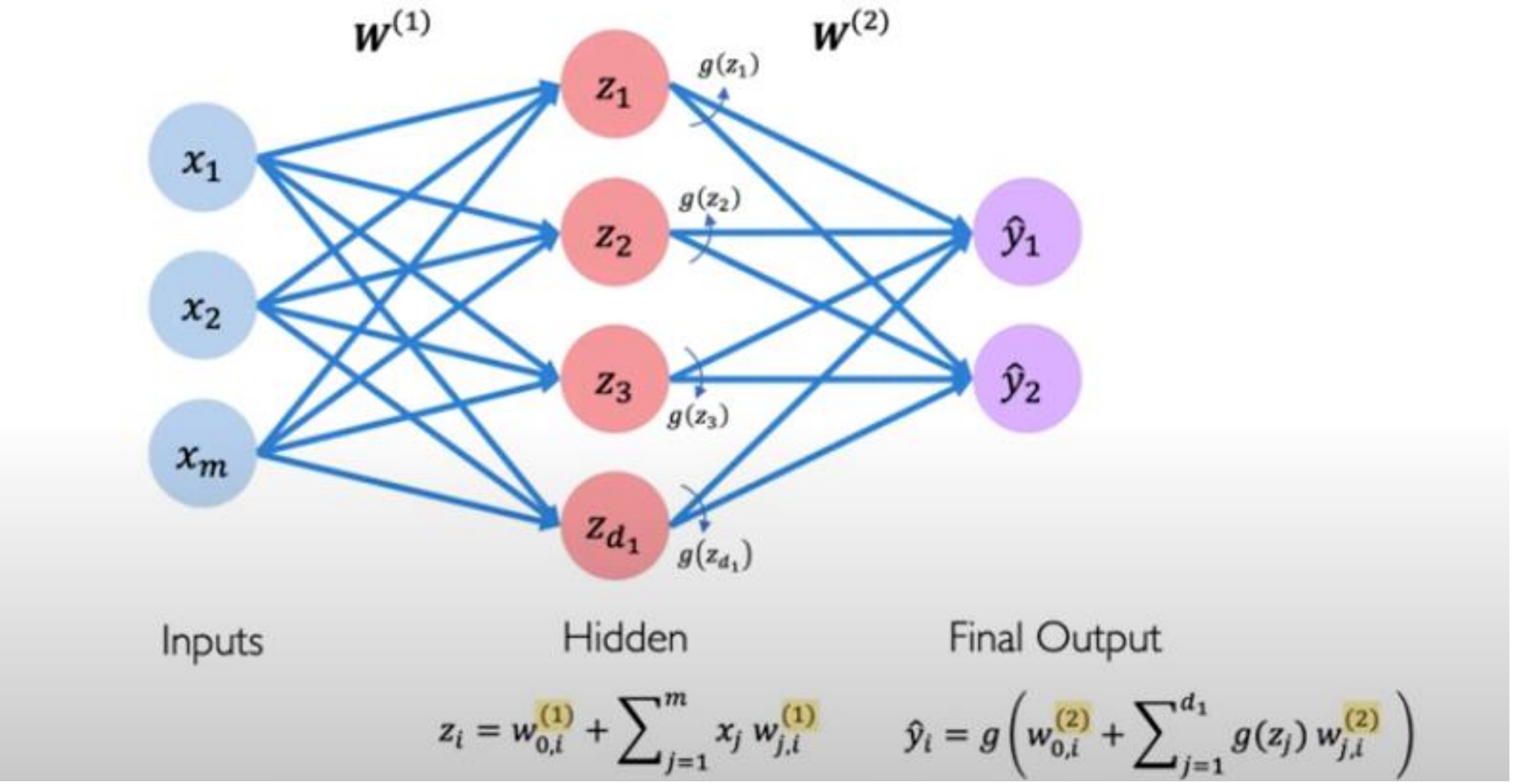
$$z = w_0 + \sum_{j=1}^{m} x_j \, w_j$$

1. We can further simplified the equation and diagram as above. The bias term w_0
2. Apply the function. In this case the function is a nonlinear function such as sigmoid function

Revite: Multiple Output Perceptron

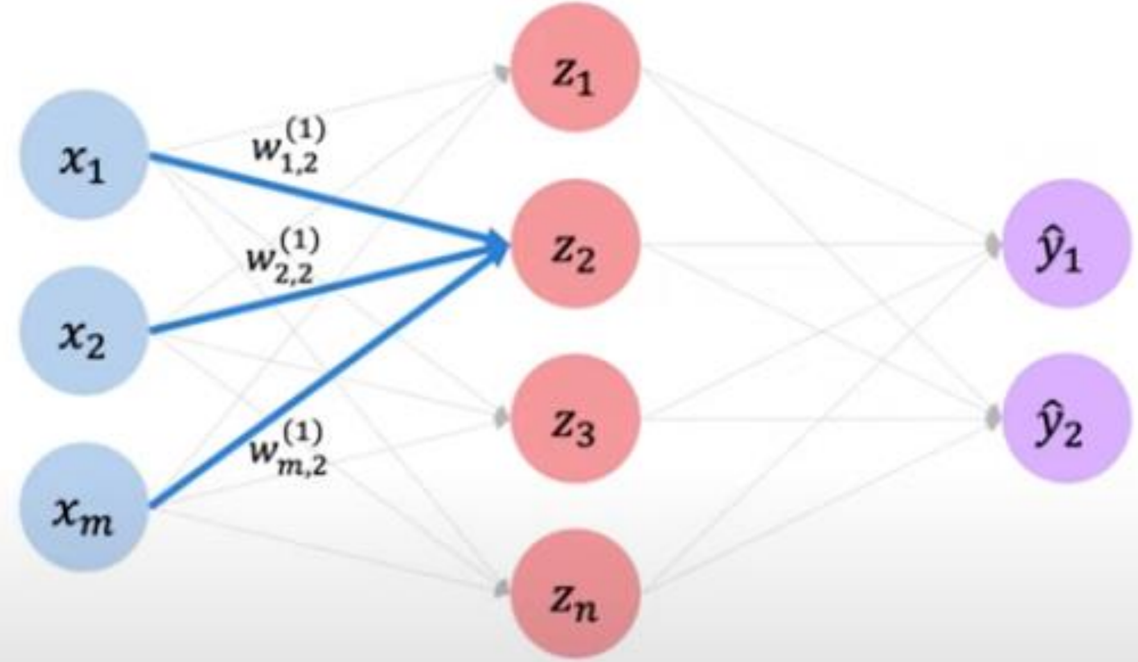


Revisit: A single hidden layer ANN



1. The layer in between the input and output layer is called the hidden layer
2. The state of the hidden layers are typically unobserved or hidden.

Revisit: The single layer ANN

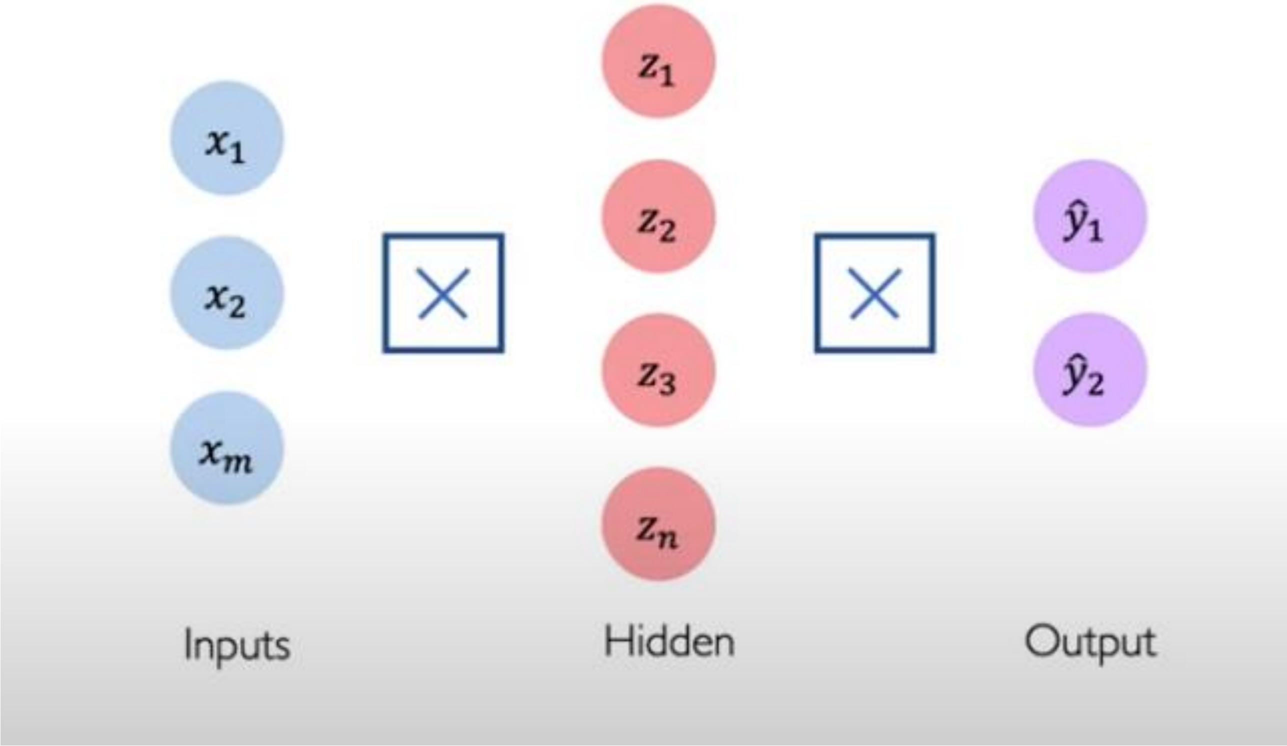


$$z_2 = w_{0,2}^{(1)} + \sum_{j=1}^{m} x_j\, w_{j,2}^{(1)}$$

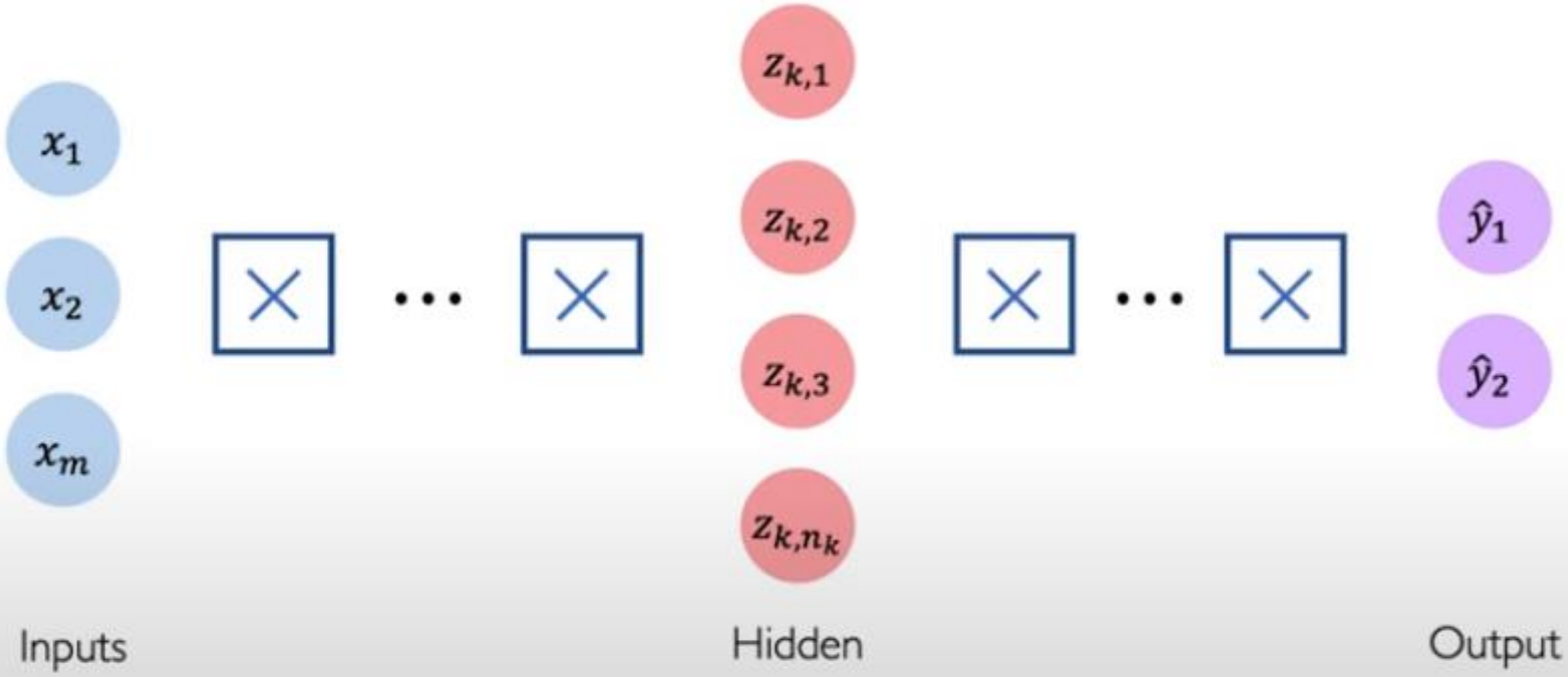
$$= w_{0,2}^{(1)} + x_1\, w_{1,2}^{(1)} + x_2\, w_{2,2}^{(1)} + x_m\, w_{m,2}^{(1)}$$

An example of computing the hidden layer second neuron (Z_2)

How do we extend too multiple hidden layers?

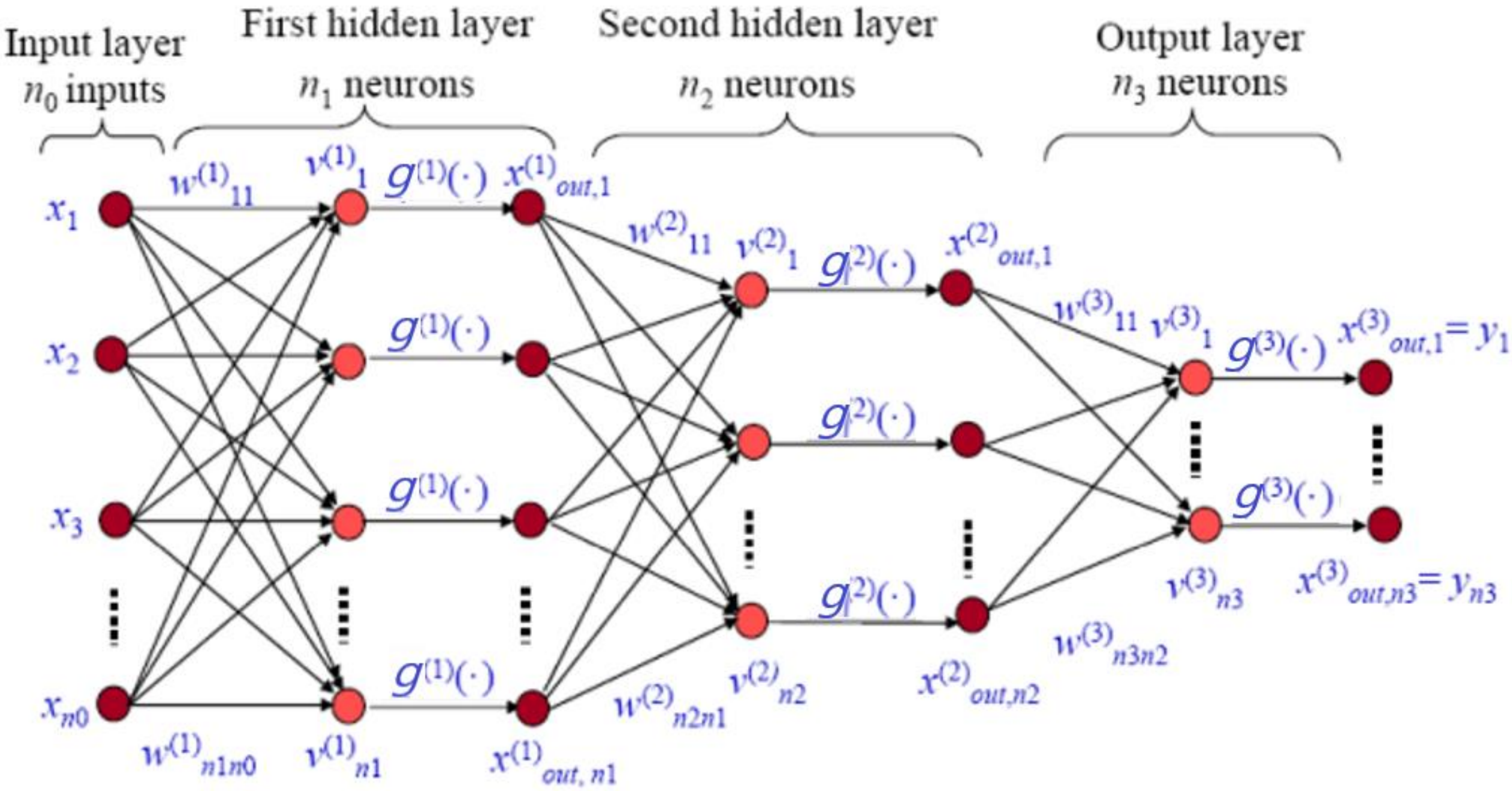


Multi-layers ANN or Deep Neural networks



$$z_{k,i} = w_{0,i}^{(k)} + \sum_{j=1}^{n_{k-1}} g(z_{k-1,j})\, w_{j,i}^{(k)}$$

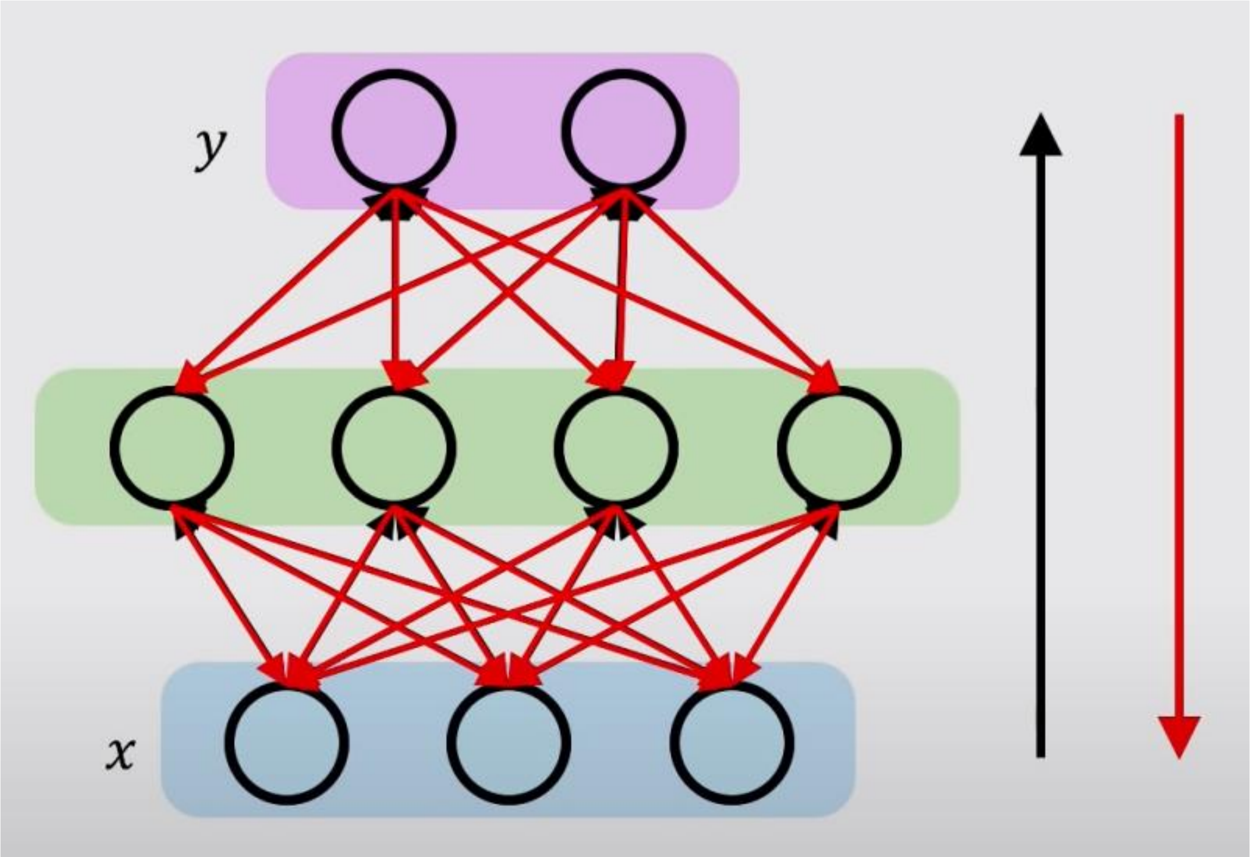
Example of Multi-Layer Perceptron (MLPs)



Reference from Neural Network EE5904 Lecture 3

Multi-Layer Perceptron (MLPs)

- Repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector.



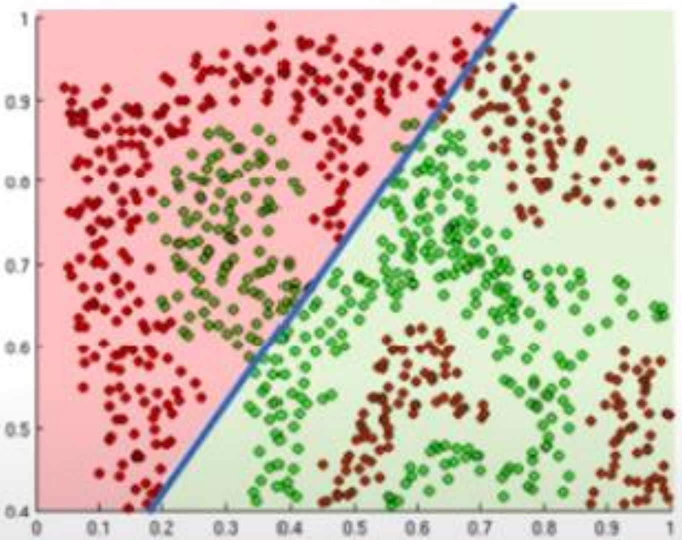
https://www.youtube.com/watch?v=SEnXr6v2ifU&list=RDCMUCtslD4DGH6PKyG_1gFAX7sg&index=3

Backpropagation process or algorithm:

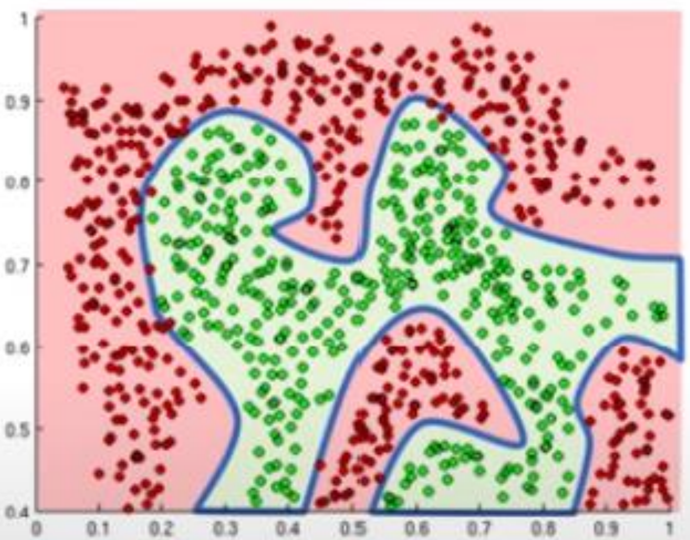
1. Take the derivative (gradient) of the loss with respect to each parameter
2. Update/adjust the parameters value in order to minimize loss

Importance of Activation Functions

The purpose of activation functions is to ***introduce non-linearities*** *into the network*



Linear function activation function produce linear decision



Nonlinear activation function allow us to approximate arbitrarily complex function

Reference: MITIntroduction to Deep Learning | 6.S191

What is an Activation Function?

National University of Singapore

An activation function in ANN defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network.

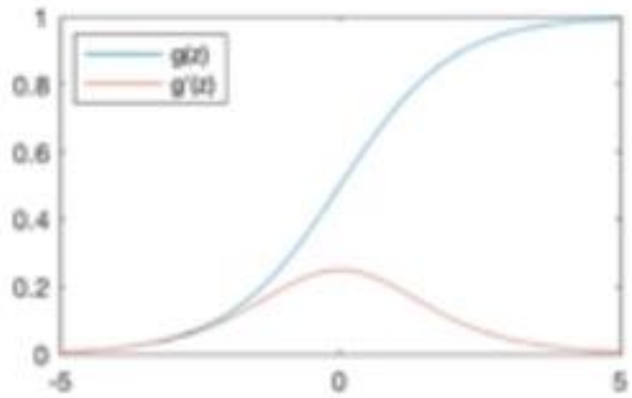
Activation function is called a "Transfer function" or "Squashing function". Squashing as the output range of the activation function is limited or being squashed. Many activation functions are nonlinear and may be referred to as the "nonlinearity" in the layer or the network design.

The choice of activation function has a large impact on the capability and performance of the ANN, and different activation functions may be used in different parts of the model; depending on the design. However, in most general cases, the same nonlinear activation function can be used throughout the hidden layers of the ANN design.

Common Activation Functions (g or f) for ANN

NUS

Sigmoid Function

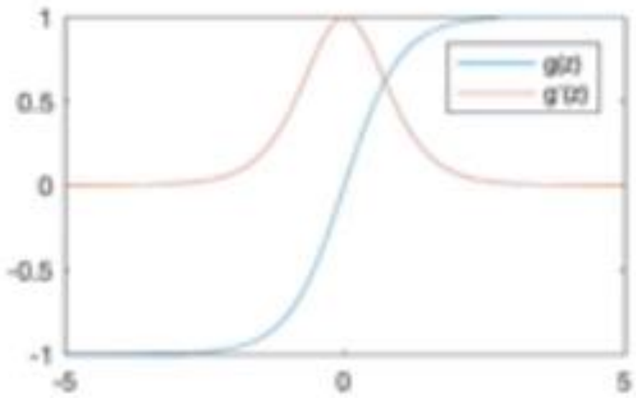


$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g'(z) = g(z)(1 - g(z))$$

```
tf.math.sigmoid(z)
```

Hyperbolic Tangent

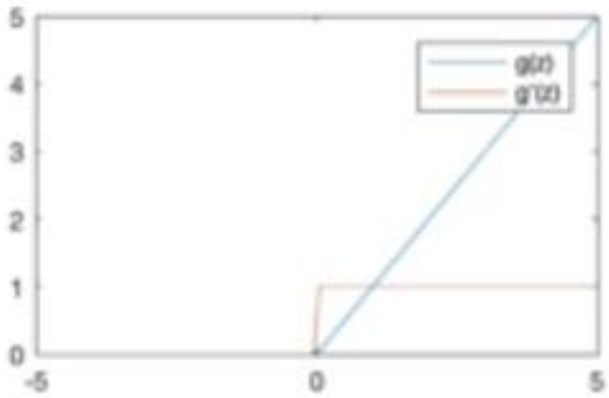


$$g(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

$$g'(z) = 1 - g(z)^2$$

```
tf.math.tanh(z)
```

Rectified Linear Unit (ReLU)

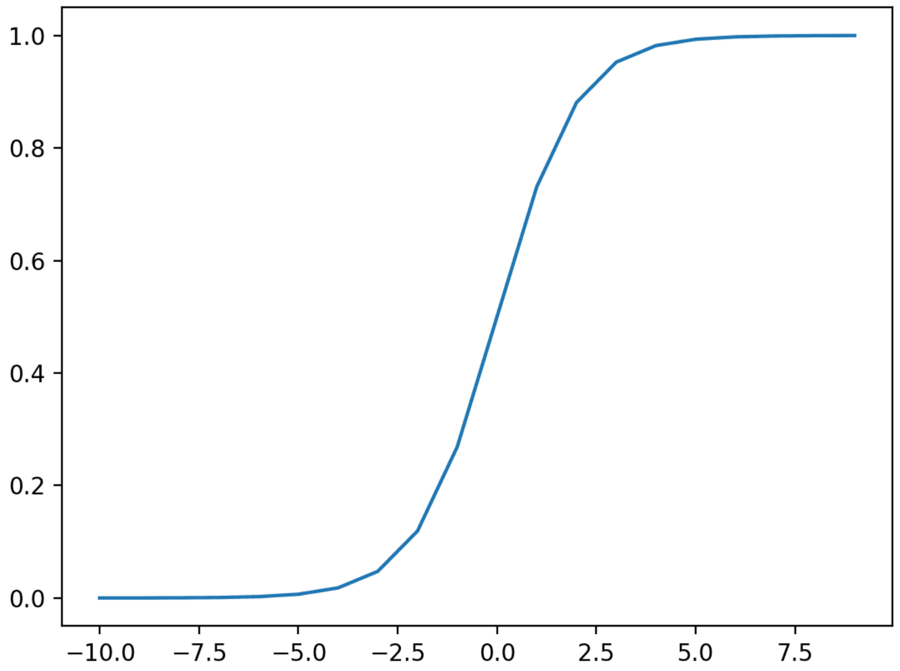


$$g(z) = \max(0, z)$$

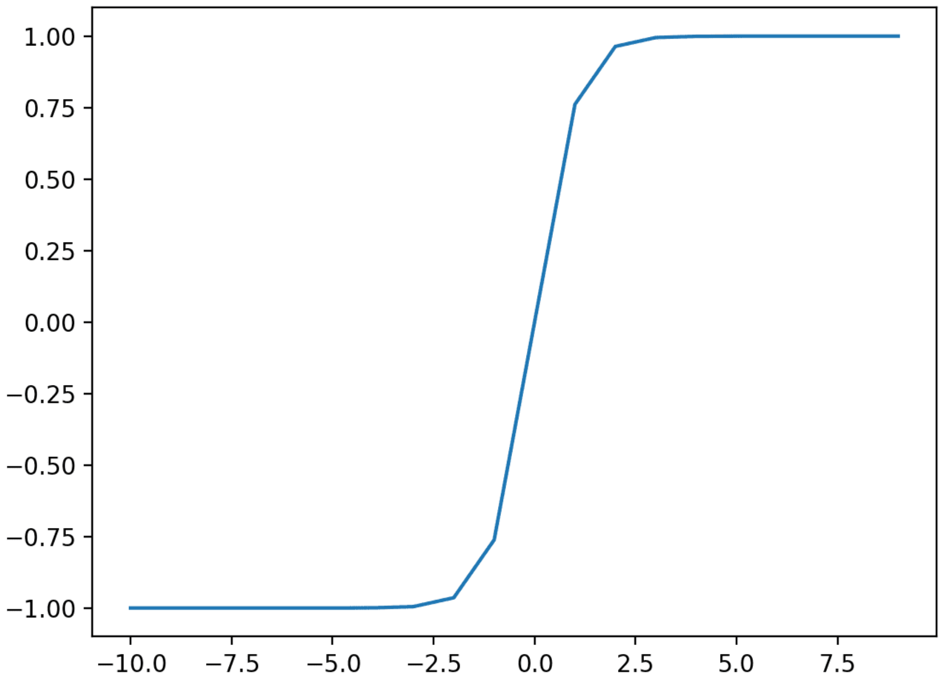
$$g'(z) = \begin{cases} 1, & z > 0 \\ 0, & \text{otherwise} \end{cases}$$

```
tf.nn.relu(z)
```

Common Activation Functions (g or f) for ANN



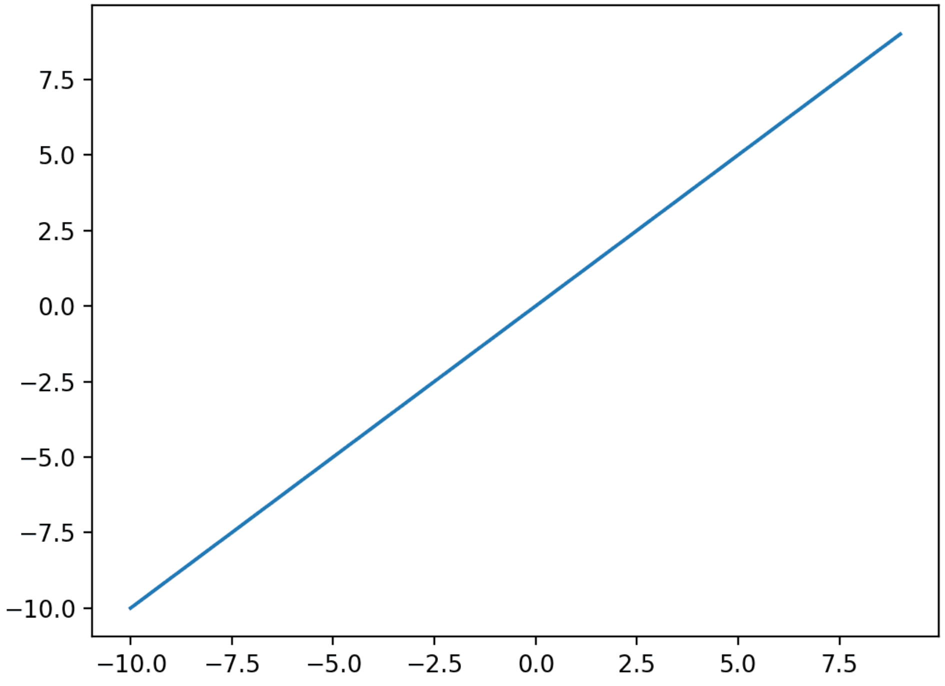
Sigmoid Function



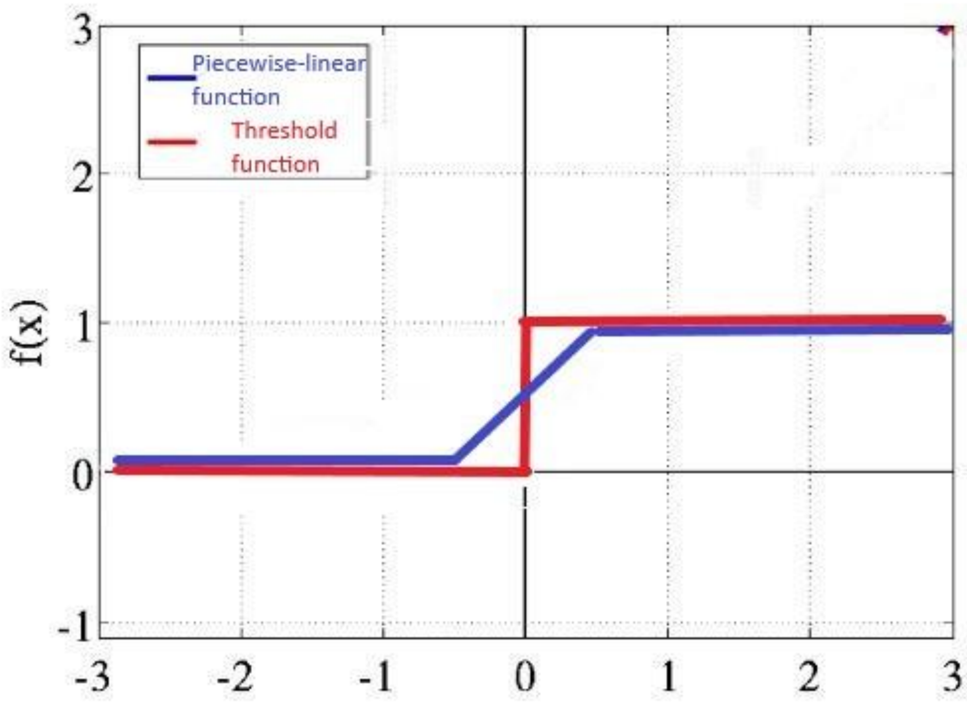
Tanh Function

Typically use at the hidden layer activation function. Tanh is more flexible as it can address both the negative and positive region of a space.

Activation Functions (*g* or *f*) of ANN



Linear Function: Typically used at the output of the ANN model



Threshold function $f(z) = \begin{cases} +1 & if\ z \geq 0 \\ 0 & if\ z < 0 \end{cases}$

Piecewise-linear function:

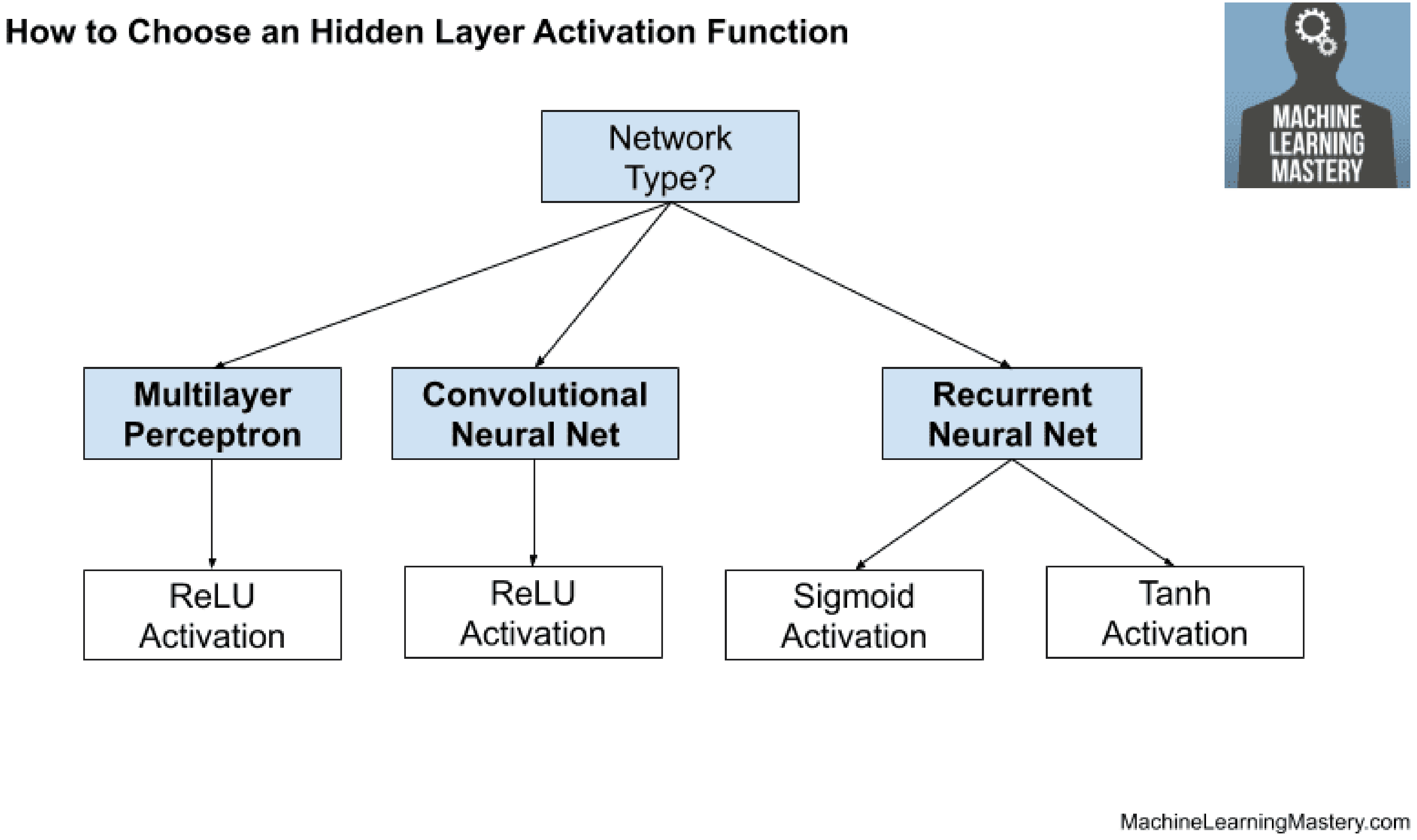
$$f(z) = \begin{cases} 1 & if\ z \geq 0.5 \\ z+0.5 & -0.5 < z < 0.5 \\ 0 & if\ z \leq -0.5 \end{cases}$$

Activation Function

Typically you need to perform first-order derivative on the activation functions. Hence knowing how to do the differentiation or derivative is importance. This is required given that neural networks are typically trained using the backpropagation of error algorithm that requires the derivative of prediction error in order to update the weights of the model.

Some exercises: Perform the first-order derivation of the common activation function such as sigmoid function and hyperbolic tangent function.

This is one approach of choosing activation function for the ANN model



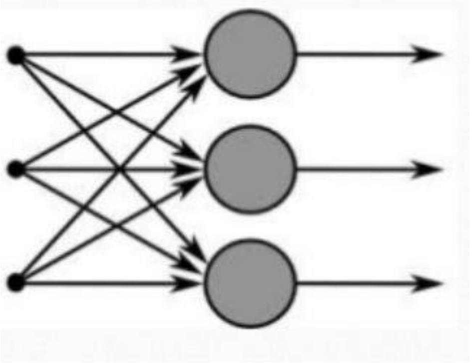
Different Types of Layered Feedforward Networks

Single-Layer Feed-forward Networks

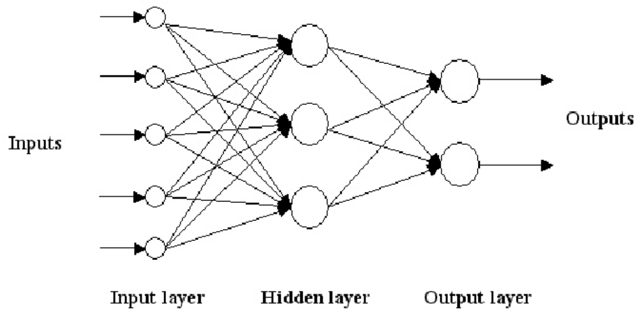
- Only one output later of computation nodes (neurons).
- Simplest model of neural network.
- Input and output are linked with each other.
- Inputs are connected to the processing nodes with various weights, resulting in series of outputs one per node.

Multilayer Feed-forward Networks

- Neurons in each layer of the network have as their inputs the output signals of the preceding layer only
- Most commonly used
- There is one or more hidden layers whose computation nodes called hidden neurons.
- Function of hidden neurons is to intervene between the external input and network output in some useful manner.



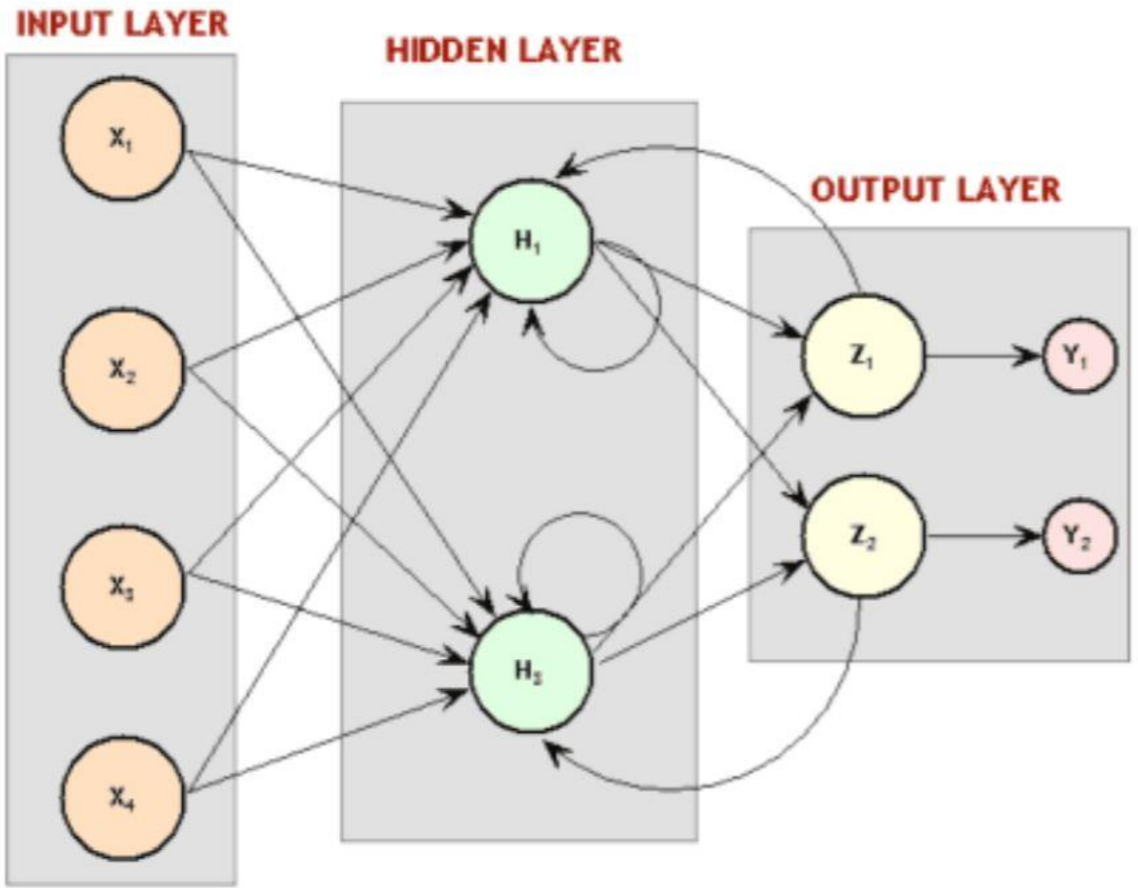
https://www.researchgate.net/profile/Murat_Sazli/publication/228394623/figure/fig2/AS:302007991717904@1449015724170/A-multi-layer-feed-forward-neural-network.png



Recurrent neural networks

Feedforward neural networks extended to include feedback connections

Example:



ADVANTAGES

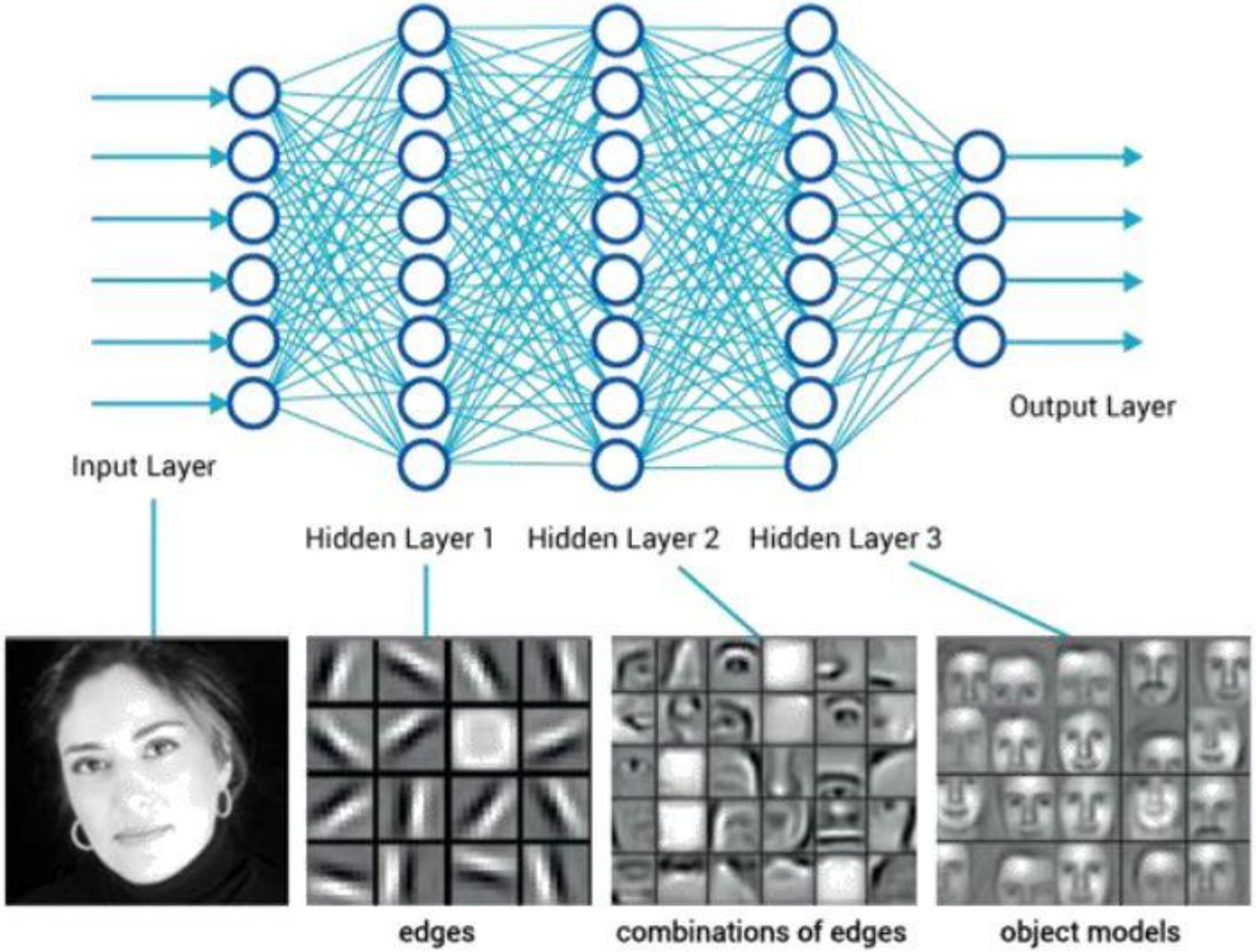
- Ability to process input of any length.
- Model size does not increase with the size of input.
- Computation takes into account of historical information.
- Weights are shared across time.

DISADVANTAGES

- Slow computation.
- Difficult to access information from a long time ago.
- Cannot consider any future input for the current state

Application of ANN in Pattern Recognition

Example: Facial Recognition

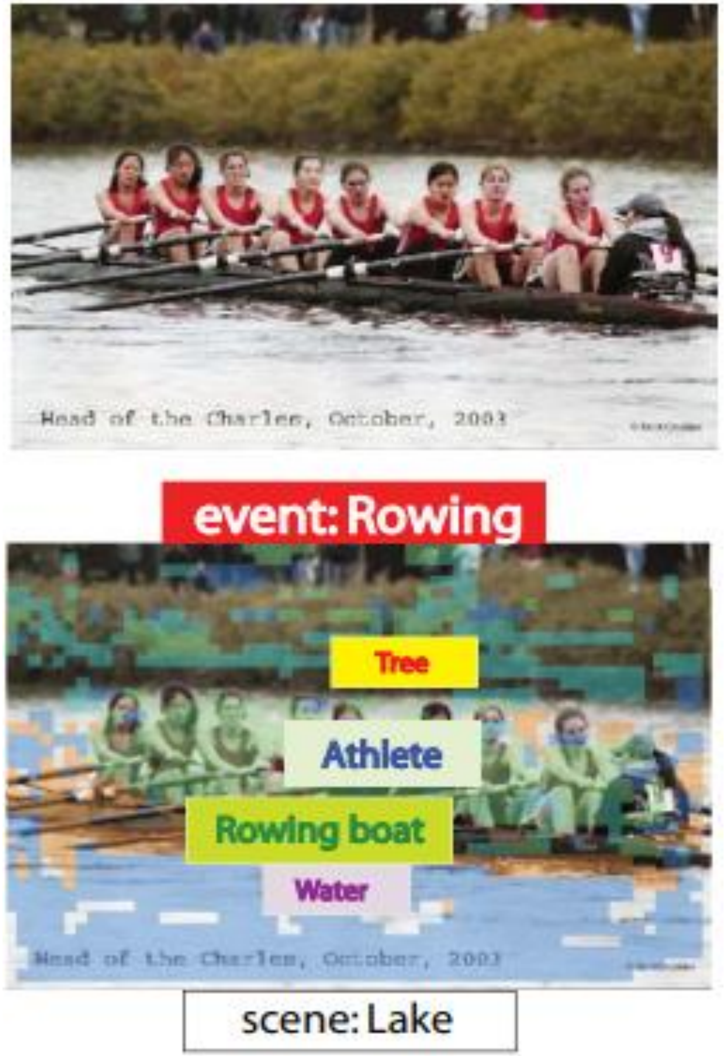


Deeper layers learn features of higher-order.

Lee, Honglak , et al. "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations." ICML

Classifying events by scene and object recognition

Classify events in static images by integrating scene and object categorizations

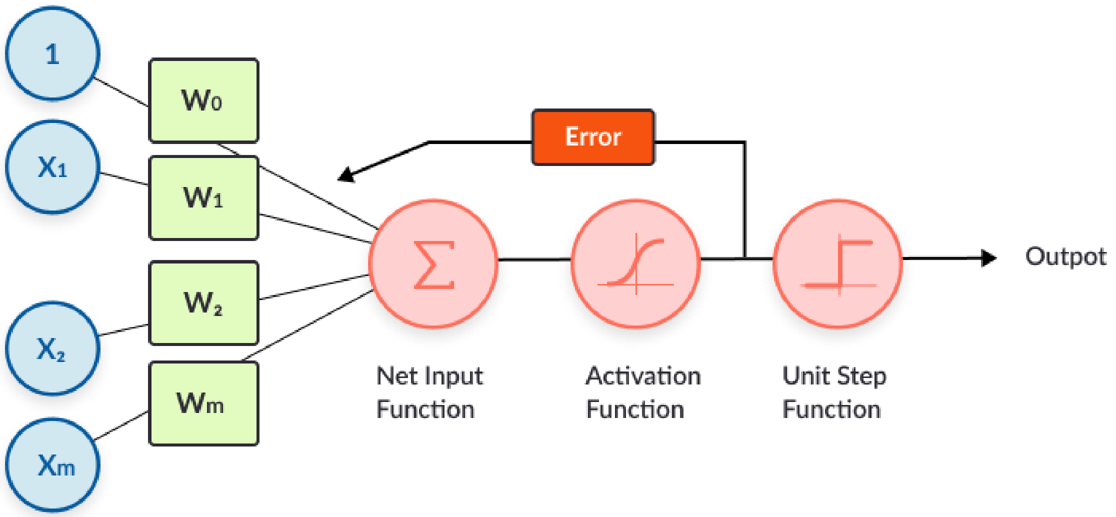


For example, given a rowing scene, the algorithm used recognizes the event as rowing by classifying the environment as a lake and recognizing the critical objects in the image as athletes, rowing boat, water, etc.

L.-J. Li and L. Fei-Fei, "What, where and who? Classifying events by scene and object recognition," 2007 IEEE 11th International Conference o
Computer Vision, 2007.c

Regression in Neural Networks

Neural networks are reducible to regression models(a neural network can "pretend" to be any type of regression model).



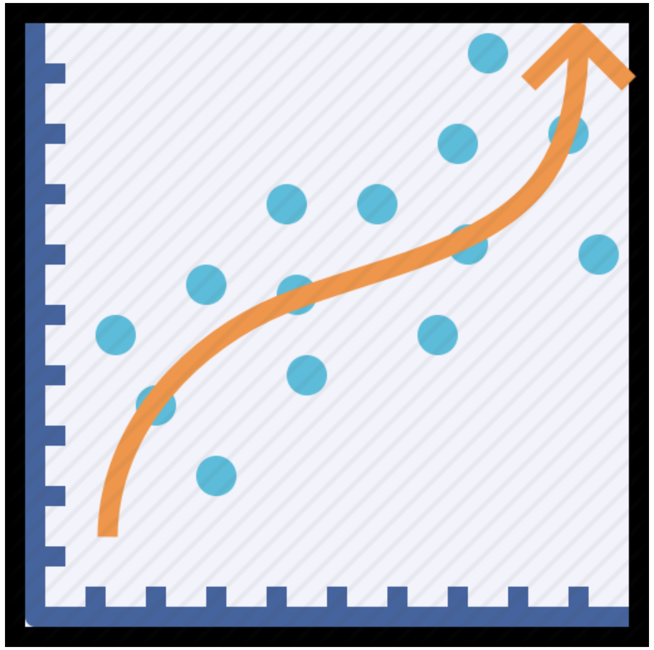
Regression in neural networks

When this neural network is trained, it will perform gradient descent to find coefficients that are better and fit the data, until it arrives at the optimal linear regression coefficients (or, in neural network terms, the optimal weights for the model).

Image from:https://missinglink.ai/wp-content/uploads/2018/12/graph3.png

Regression

Regression have proven very useful in modelling real world problems and providing useful predictions for many years, both in scientific and business industries.



The simplest, linear regression equation:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \epsilon$$

where:

y - dependent variable;

$x_2, x_3 \ldots x_\mathrm{k}$ - independent variables(one or more values that the model takes as an input, using them to predict the dependent variables) ;

1,2,3…k – Coefficients (weights that define how important each of the variables is for predicting the dependent variable);

ϵ – Error(the distance between the value predicted by the model and the actual dependent variable y)

https://cdn4.iconfinder.com/data/icons/business-consultant-flat/64/trend-prediction-stat-regression-analysis-512.png

Thank you for your attention
Q&A during Tutorial Time