Big Data Technology

Paul Rad, Ph.D.

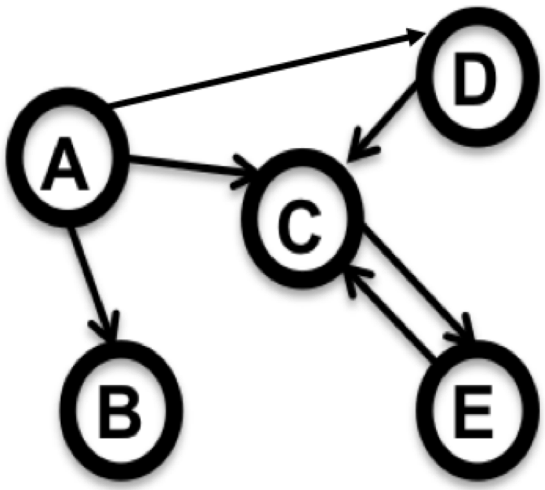
Associate Professor
Information Systems and Cyber Security, College of Business School
Electrical and Computer Engineering, College of Engineering

Outline

- **Logistics**
- **Summary and Review Previous Sections**
- **Machine Learning**

Question 1)

Draw the adjacency matrix for the following graph



To

From

	A	B	C	D	E
A					
B					
C					
D					
E					

Question #2)

A loop in a graph is where ….

There is a way to get from node A to node B and back to node A

There is some (arbitrary length) path from a node back to itself, passing through other nodes

There is an edge from a node to itself

All of the above

Question #3)

What does Betweenness Centrality mean?

1. **The average of the shortest distance to all other nodes in the graph**
2. **The number of edged connected to a node**
3. **Extent to which a particular node lies on the shortest path between other nodes**
4. **A measure of the extent to which a node is connected to influential other nodes**

Question 4

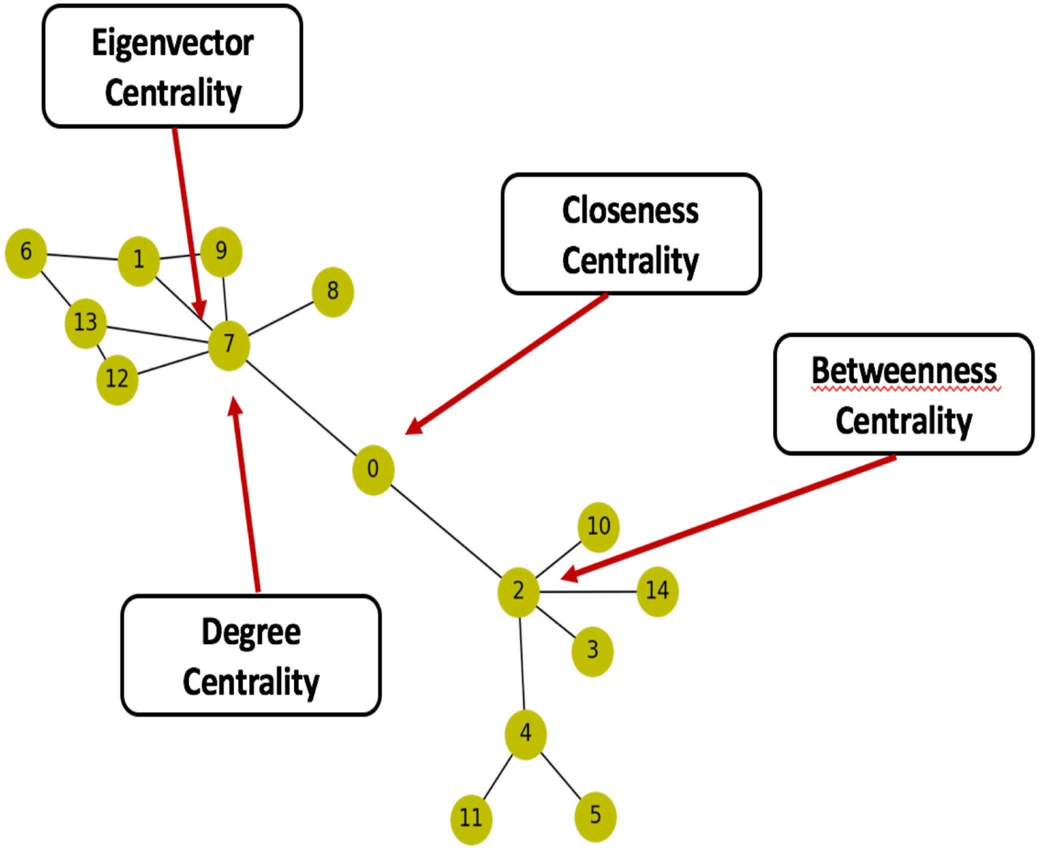
In the following graph, which node (network switch) will have the highest impact if it shuts down by cyber attackers?

Node 2, a Between centrality

Node 0, a Closeness centrality

Node 7, a Degree centrality

Node 7, an Eigenvector centrality



A Few Quotes

“A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)

“Machine learning is the next Internet”
(Tony Tether, Director, DARPA)

Machine learning is the hot new thing”
(John Hennessy, President, Stanford)

“Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)

A Short History of Machine Learning

1950 — **Alan Turing** creates the "Turing Test" to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human.

1952 — **Arthur Samuel** wrote the first computer learning program. The program was the game of checkers, and the IBM computer improved at the same the more it played studying which moves made up winning and incorporating those moves into its program.

1957 — **Frank Rosenblatt** designed the first neural network for computers (the perceptron), which simulate the thought processes of the human brain.

1967 — The "nearest neighbor" algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for traveling salesmen, starting at a random city but ensuring they visit all cities during a short tour.

1979 — Students at Stanford University invent the "Stanford Cart" which can navigate obstacles in a room on its own.

1981 — **Gerald Dejong** introduces the concept of Explanation Based Learning (EBL), in which a computer analyses training data and creates a general rule it can follow by discarding unimportant data.

1985 — **Terry Sejnowski** invents NetTalk, which learns to pronounce words the same way a baby does.

A Short History of Machine Learning

1990s — Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions — or “learn” — from the results.

1997 — IBM’s Deep Blue beats the world champion at chess.

2006 — Geoffrey Hinton coins the term “deep learning” to explain new algorithms that let computers “see” and distinguish objects and text in images and videos.

2010 — The Microsoft Kinect can track 20 human feature at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.

2011 — IBM’s Watson beats its human competitors at Jeopardy.

2011 — Google Brain is developed, and its deep neural network can learn to discover and categorize objects much the way a cat does.

2012 – Google’s X Lab develops a machine learning algorithm that is able to autonomously browse YouTube videos to identify the videos that contain cats.

A Short History of Machine Learning

2014 – Facebook develops DeepFace, a software algorithm that is able to recognize or verify individuals on photos to the same level as humans can.

2015 – Amazon launches its own machine learning platform.

2015 – Microsoft creates the Distributed Machine Learning Toolkit, which enables the efficient distribution of machine learning problems across multiple computers.

2015 – Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), sign on open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.

2016 – Google's artificial intelligence algorithm beats a professional player at the Chinese board game Go, which is considered the world's most complex board game and is many times harder than chess. The AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Go competition.

Machine Learning Definition

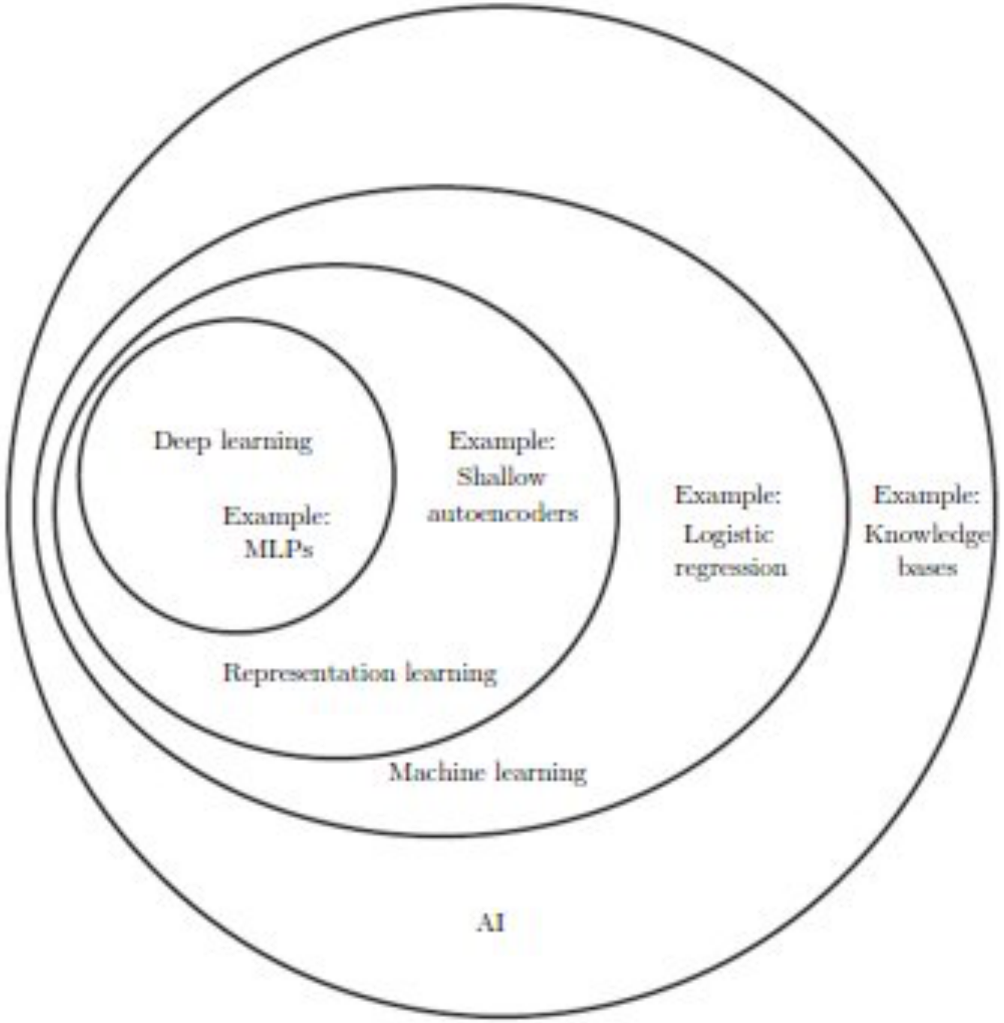
A machine learning algorithm is an algorithm that is able to learn from data.

Tom Mitchell (1997) provides the definition:

"A computer program is said to learn from

experience *E* with respect to some class of task *T* and performance measure *P*,

if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."



3 Types of Learning

Supervised

- Learning from labeled data
- E.g., Spam classification

- Classification
- Regression
- Ranking

Unsupervised

- Discover structure in unlabeled data
- E.g., Document clustering

- Clustering
- Hidden Markov Models

Reinforcement

- Learning by “doing” with delayed reward
- E.g., Chess computer

Supervised Learning

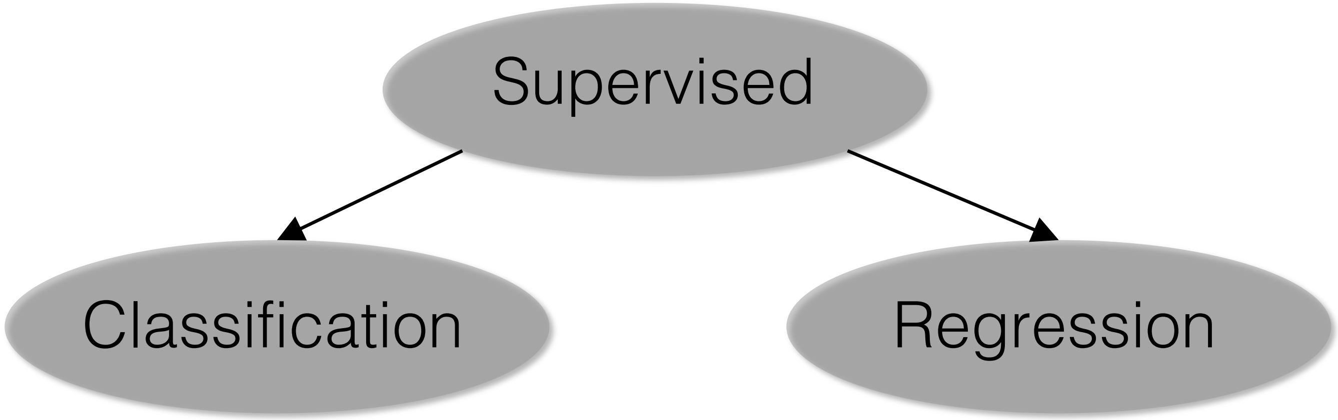
Given examples of a function *(X, F(X))*

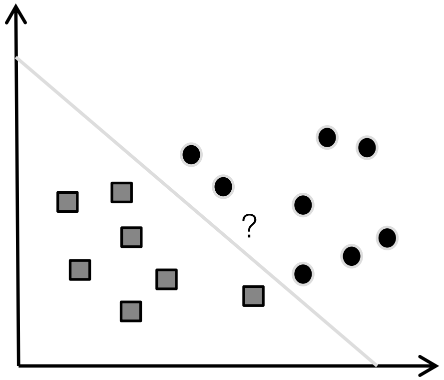
Predict function *F(X)* for new examples *X*

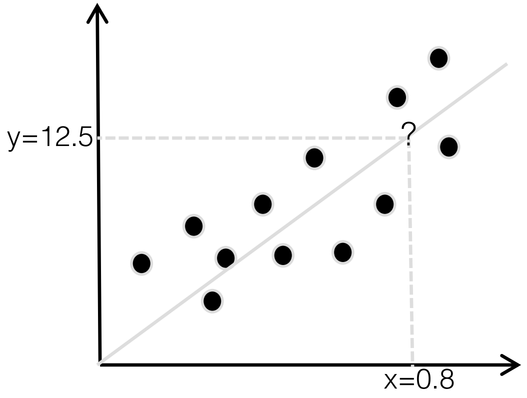
Discrete *F(X)*: Classification

Continuous *F(X)*: Regression

F(X) = Probability(*X*): Probability estimation







Regression and Classification Examples

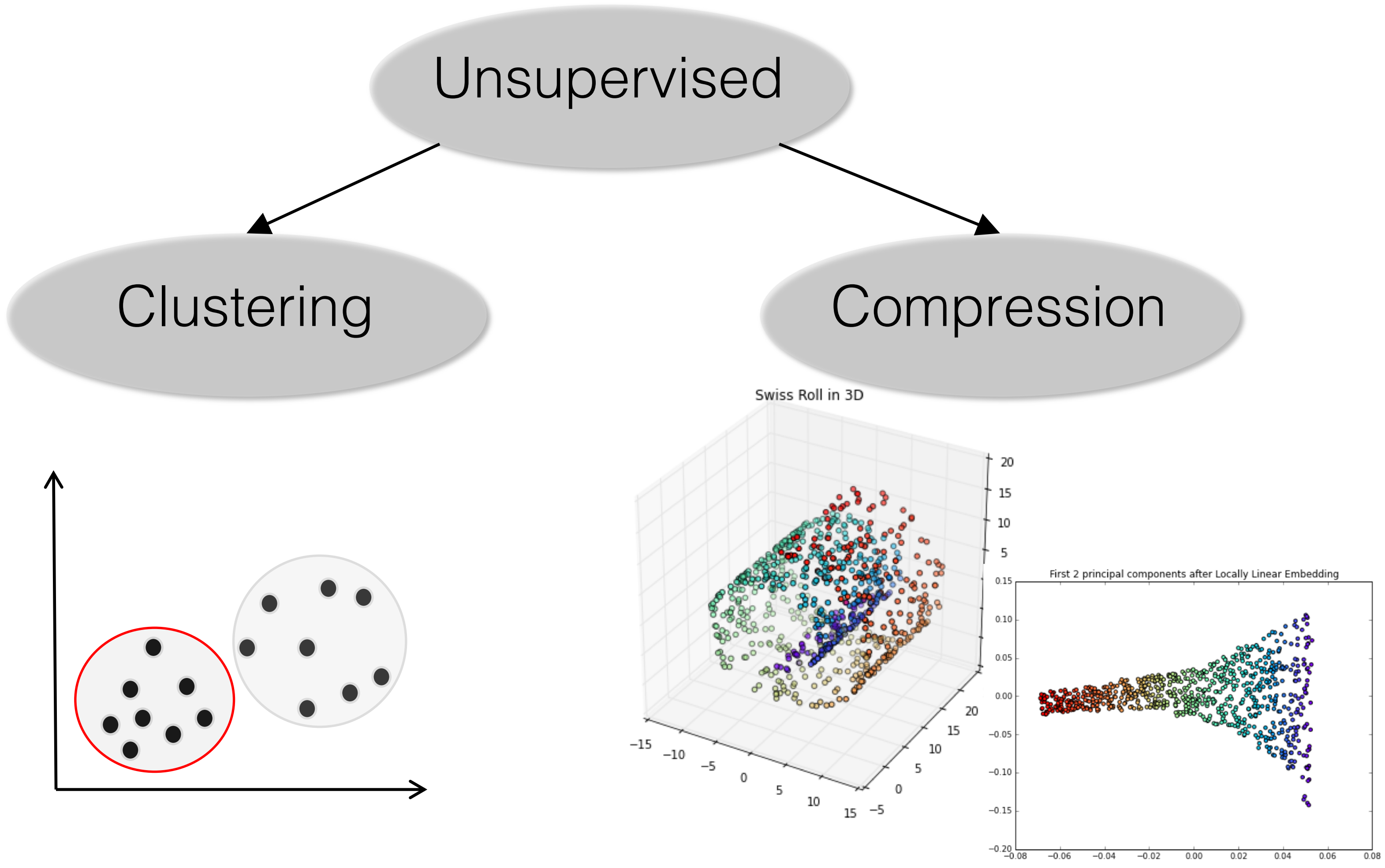
Stock prediction

* Predict the price of a stock (y)
  * Depends on x =
    - Recent history of stock price
    - News events
    - Related commodities

Spam or Not spam emails

Music or Tweeter Sentiment Analysis

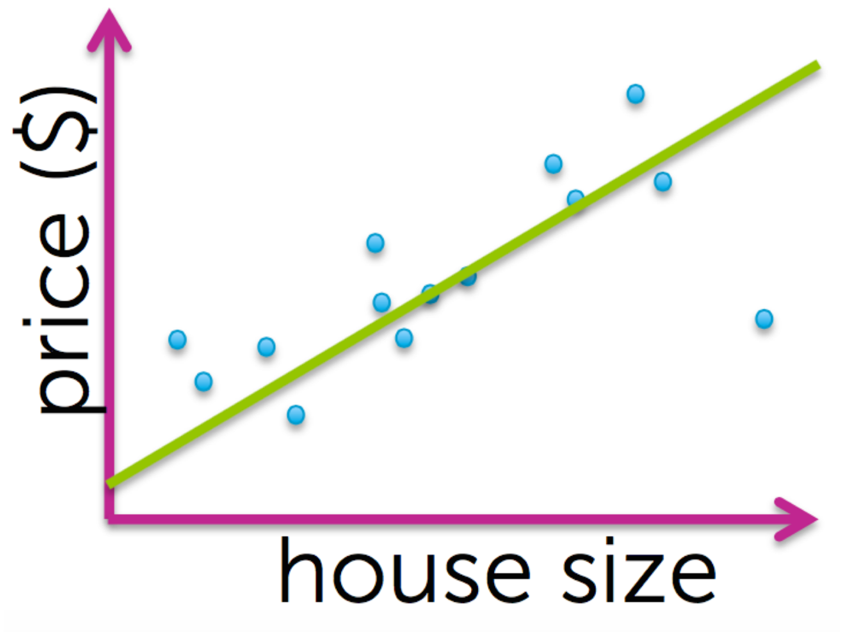
Unsupervised Learning



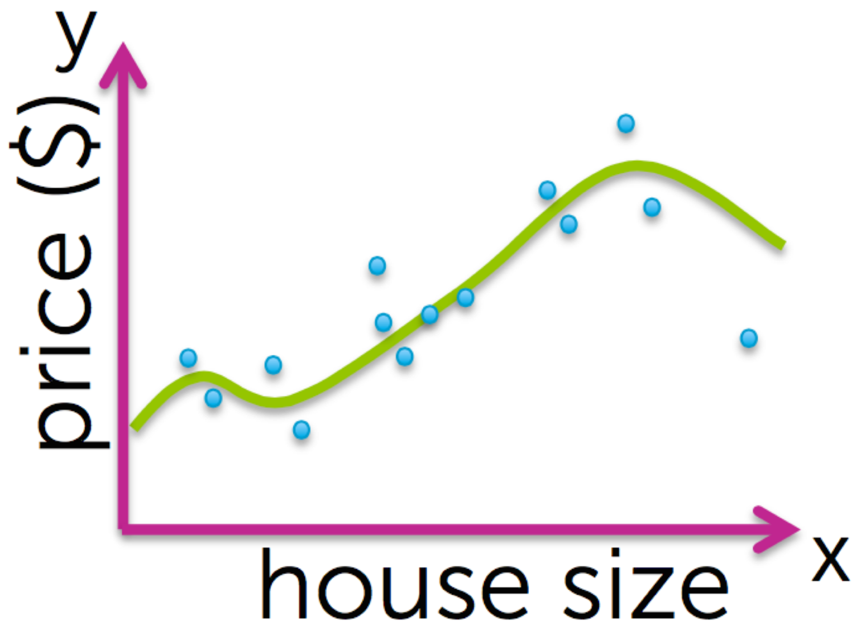
Simple Regression vs. Multiple Regression

What makes it simple?

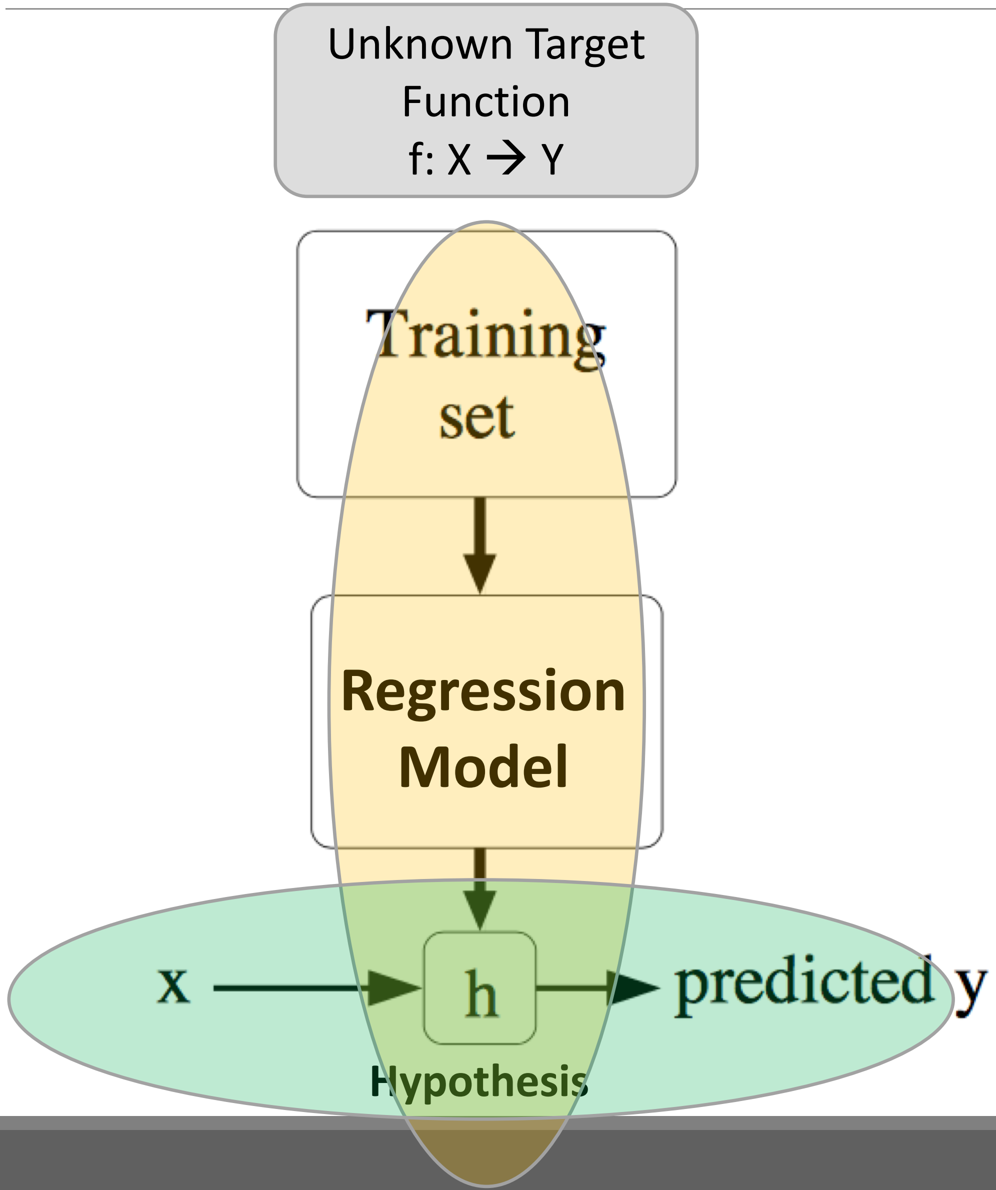
1 input and just fit a line to data



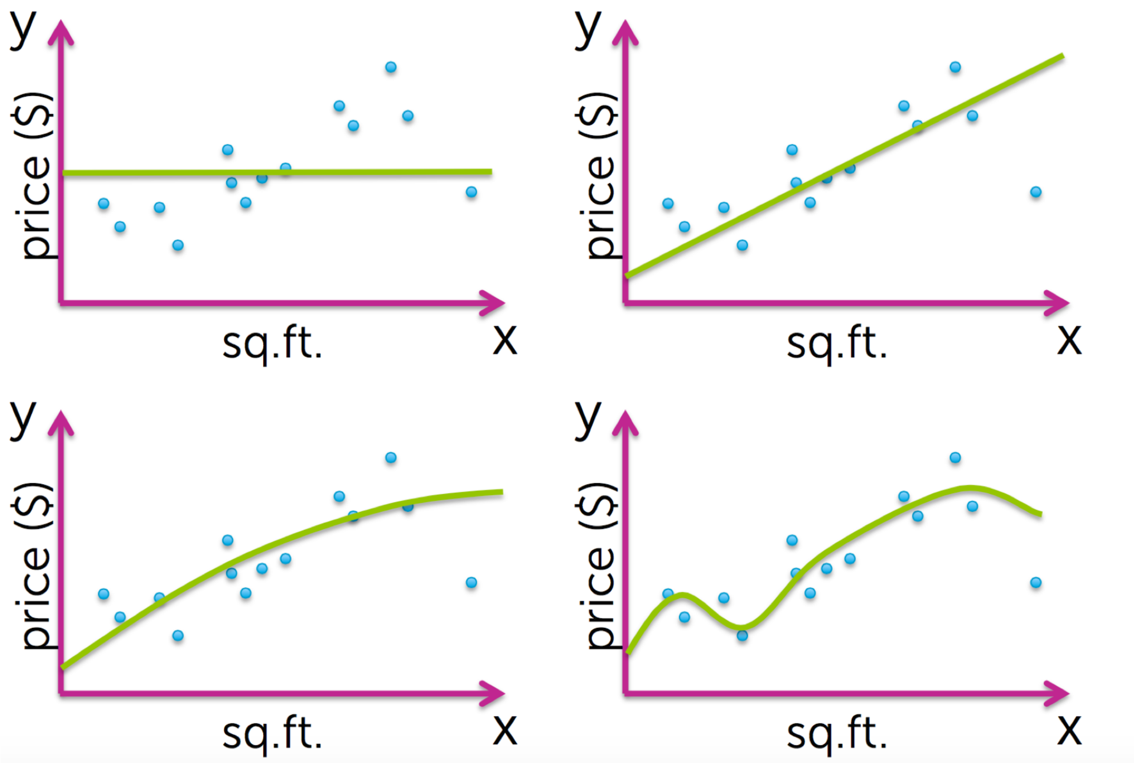
Fit more complex relationships than just a line



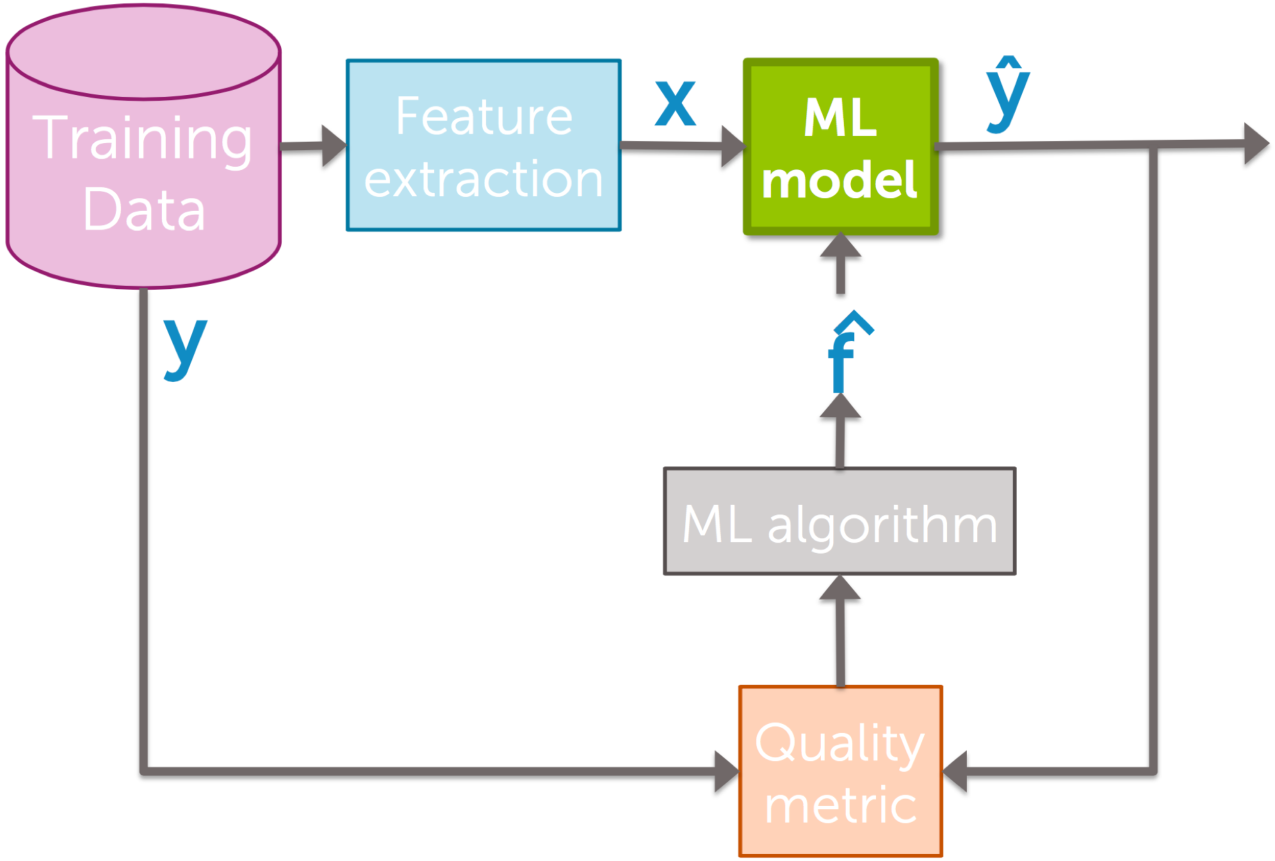
Linear Regression Learning Model

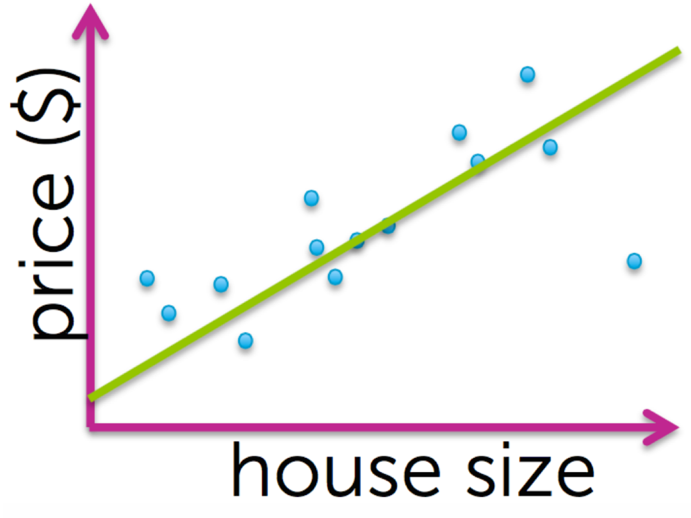


What is Linear Regression?



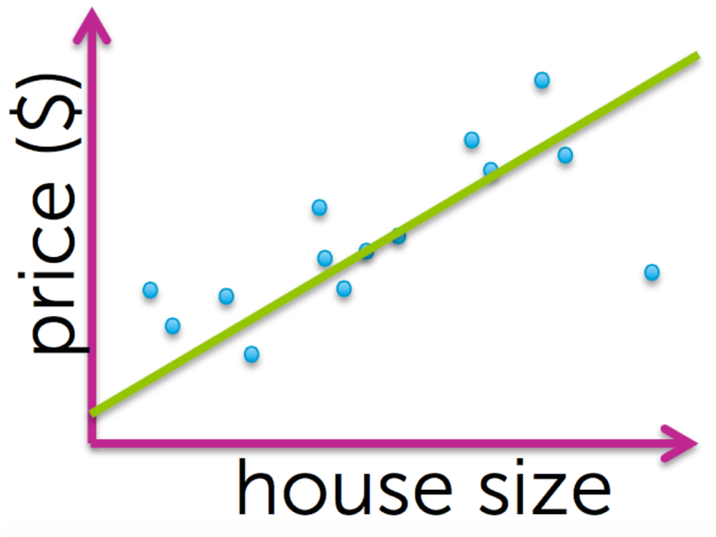
Regression Model





Predicting house prices

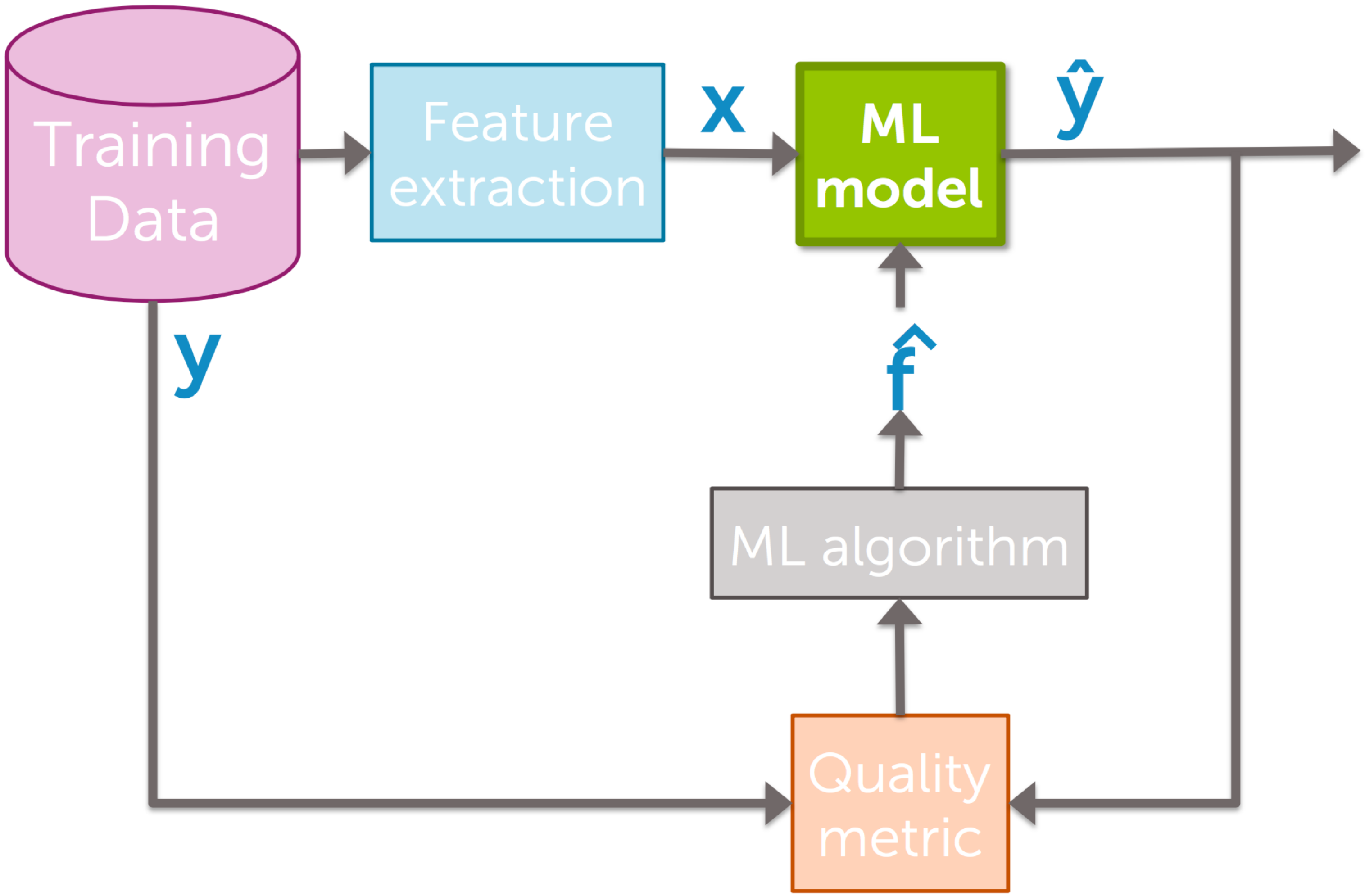
How much is my house worth?



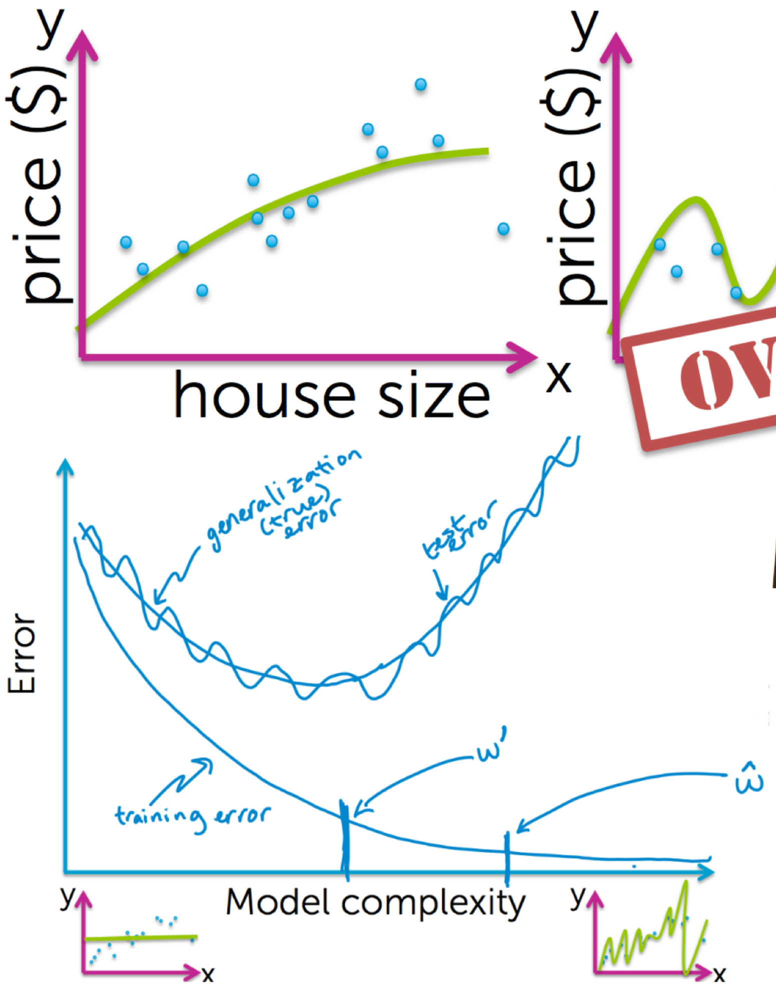
Look at recent sales in my neighborhood

- How much did they sell for?

Regression Model



Assessing Performance



Measures of error:

- Training
- Test
- True (generalization)

Neural Network History

- **Algorithms that try to mimic the brain**
- **It was very widely used in 80s and early 90s.**
- **Recently very popular again due to State-of-the-art technique for many applications. Human brain: Massively parallel network**
- **The basic computational unit of the brain is a neuron. Approximately 86 billion neurons can be found in the human nervous system and they are connected with approximately 10^14 - 10^15 synapses**

What is Neural Network or Deep Learning useful for?

Like all machine learning: making predictions

Deep learning does it with higher accuracy for some problems

For example:

- Self-driving cars
- Stock market prediction
- Recognizing faces in an image

Under active development and research

The Brain as a Universal Learning Machine

Two viewpoints on the Mind:

1) Evolutionary psychologists propose that the mind is made up of genetically influenced and domain-specific mental algorithms or computational modules, designed to solve specific evolutionary problems of the past

An evolutionary perspective leads one to view the mind as a **crowded zoo of evolved**, **domain-specific programs**. Each is functionally specialized for solving a different adaptive problem that arose during hominid evolutionary history, such as face recognition, foraging, mate choice, heart rate regulation, sleep management, or predator vigilance, and each is activated by a different set of cues from the environment.

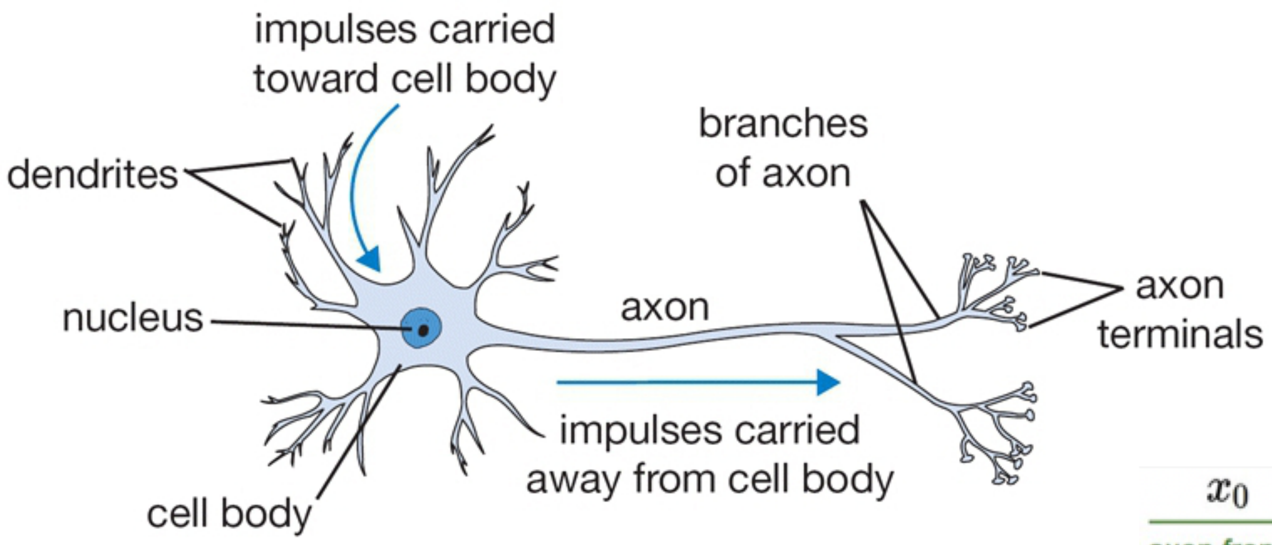
There is another viewpoint cluster, more popular in computational neuroscience (especially today), that is almost the *exact opposite* of the evolved modularity hypothesis.

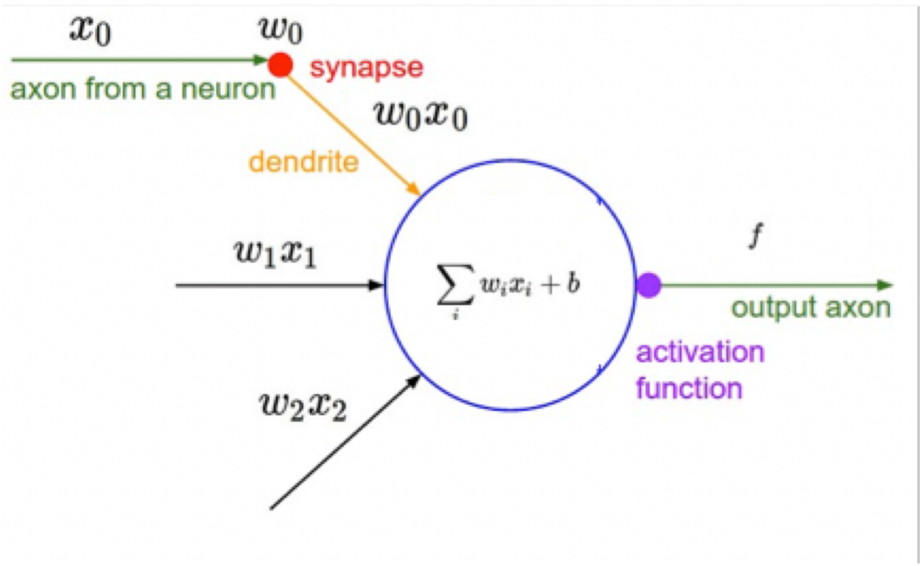
2) Universal learner" hypothesis, aka the "one learning algorithm" hypothesis

The universal learning hypothesis proposes that *all* significant mental algorithms are learned; nothing is innate except for the learning and reward machinery itself (which is somewhat complicated, involving a number of systems and mechanisms), the initial rough architecture (equivalent to a prior over mindspace), and a small library of simple innate circuits (analogous to the operating system layer in a computer). In this view the mind (software) is distinct from the brain (hardware). The mind is a complex software system built out of a general learning mechanism.

https://youtu.be/r9mvRRwu5Gw

Biological neuron and mathematical model

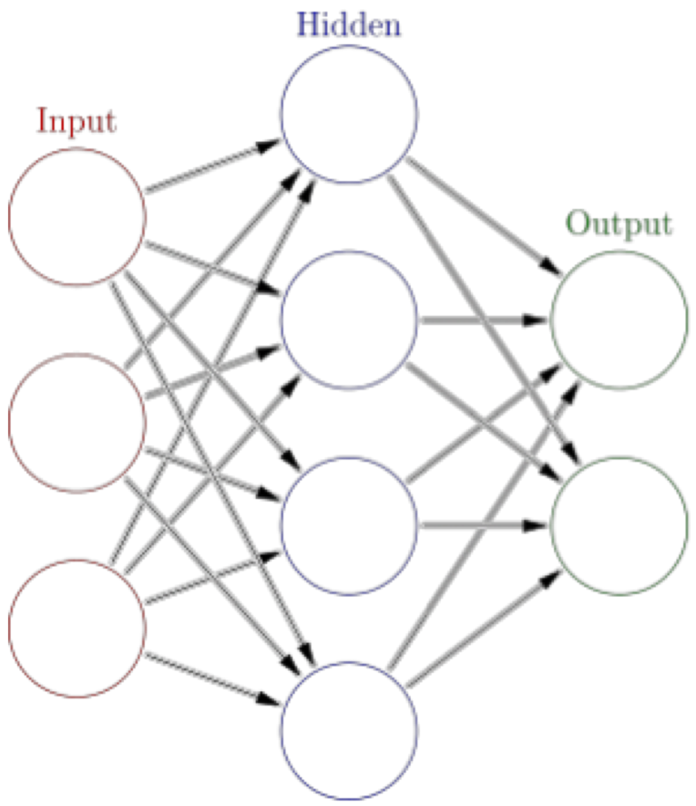




What Is A Neural Network?

The simplest definition of a neural network, more properly referred to as an 'artificial' neural network (ANN), is provided by the inventor of one of the first neurocomputers, Dr. Robert Hecht-Nielsen. He defines a neural network as:

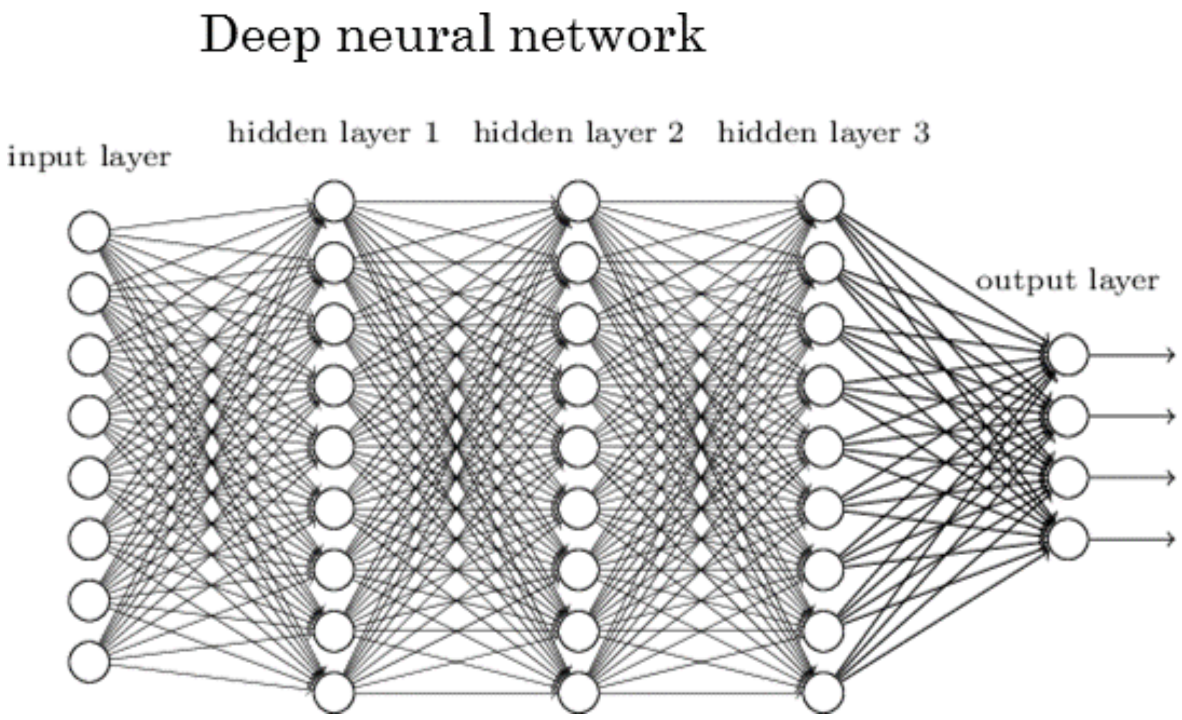
"...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. - In "Neural Network Primer: Part I" by Maureen Caudill, AI Expert, Feb. 1989



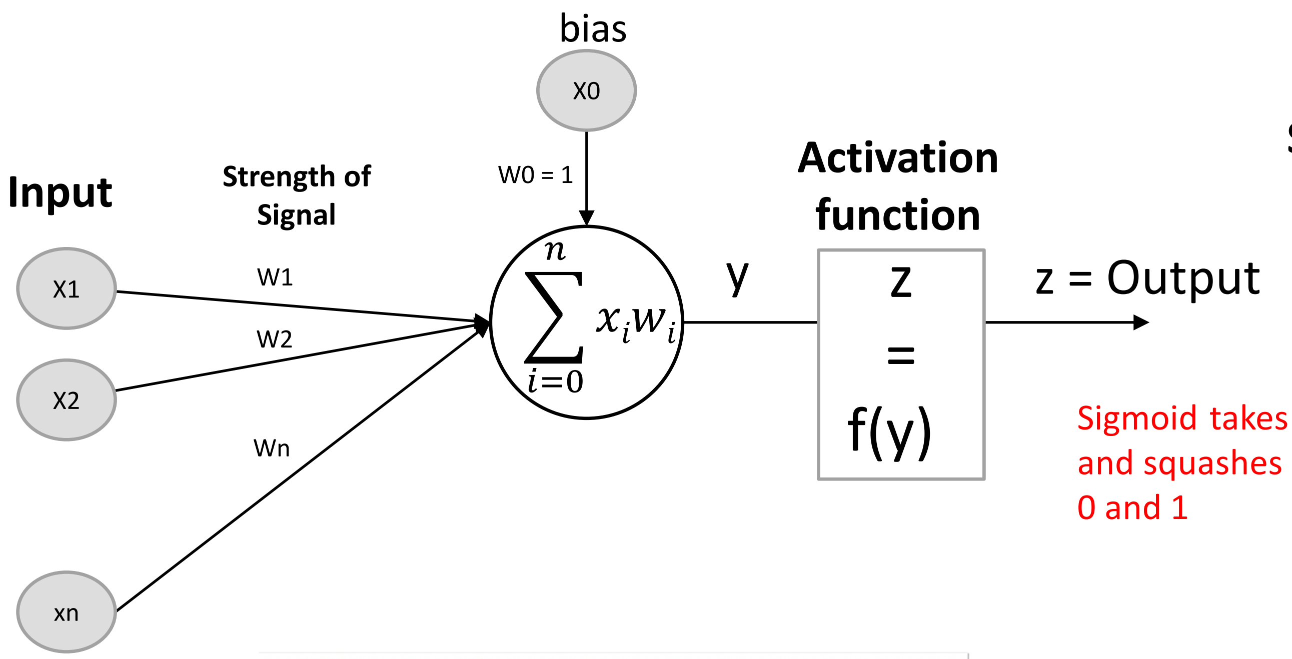
- **Every node in one layer is connected to every node in the next layer.**
- **Signals get transmitted from the input, to the hidden layer, to the output**

Deep Learning Architecture

- **An Artificial Neural Network with one or more hidden layers = Deep Learning**
- **They typically consist of many hundreds of simple processing units which are wired together in a complex communication network.**
- **Each unit or node is a simplified model of a real neuron which fires (sends off a new signal) if it receives a sufficiently strong input signal from the other nodes to which it is connected.**
- **The output is aiming for a target.**



Single Layer Neural Networks with Nonlinear Math Model

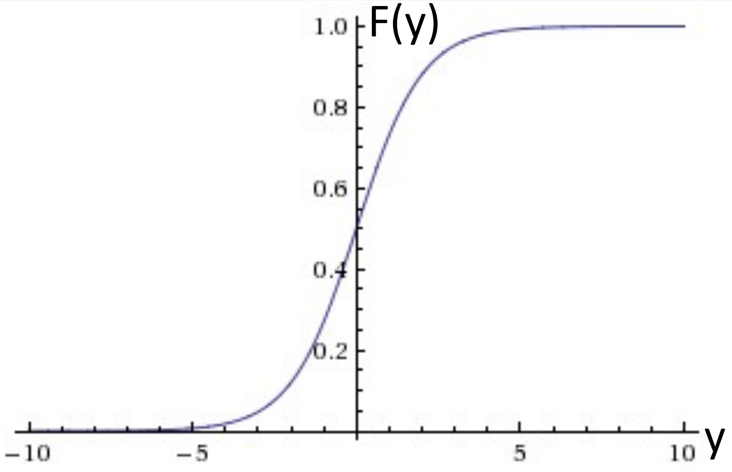


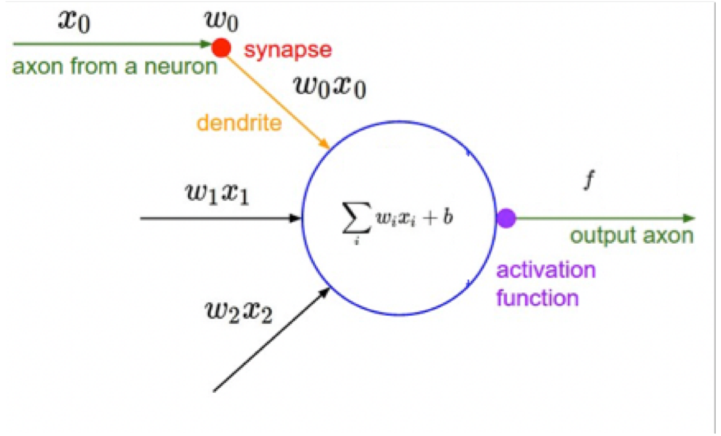
Sigmoid function:

$$f(y) = \frac{1}{1+e^{-y}}$$

$$\left(0 , 1 \right)$$

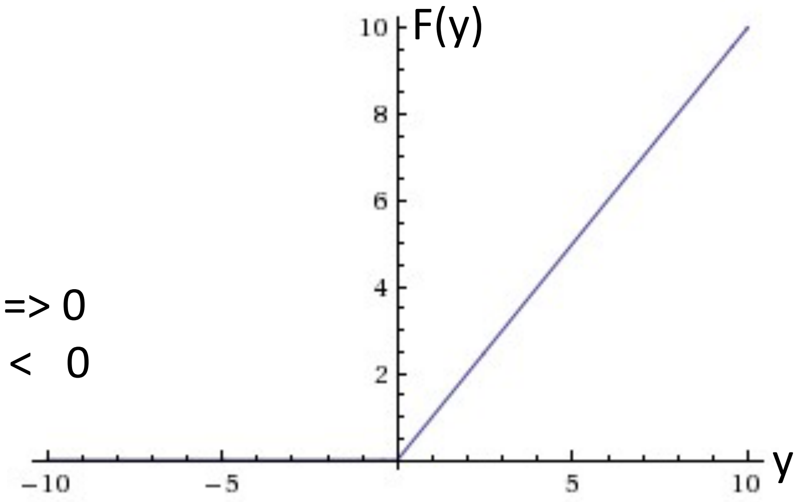
Sigmoid takes a real-valued input and squashes it to range between 0 and 1





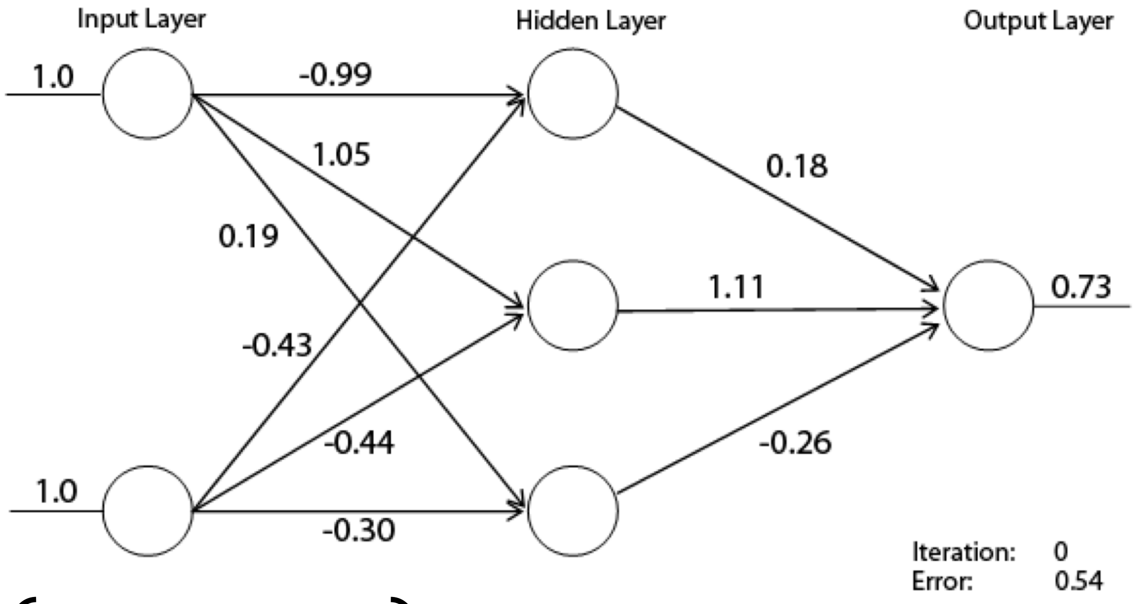
ReLU function:

$$f(y) = \begin{cases} y & , y => 0 \\ 0 & , y < 0 \end{cases}$$



Example

X1	X2	Output (XOR)
1.0	1.0	0.0



$$\begin{pmatrix} 1.0. & 1.0 \end{pmatrix} \begin{pmatrix} -0.99 & 1.05 & 0.19 \\ -0.43. & -0.44 & -0.30 \end{pmatrix} = \begin{pmatrix} 0.56 & 0.61 & -0.11 \end{pmatrix}$$

$$\begin{pmatrix} 0.56 & 0.61 & -0.11 \end{pmatrix} \begin{pmatrix} 0.18 \\ 1.11 \\ -0.26 \end{pmatrix} = \begin{pmatrix} 0.749 \end{pmatrix}$$

Y = 0.0
Y-hat = 0.749

Error = 0.749 – 0 = 0.749

Backpropagation Algorithm

Initialize all weights $w_{ij}^{(l)}$ at random

for $t = 0, 1, 2, \ldots$ do

  Pick $n \in \{1, 2, \cdots, N\}$

  Forward: Compute all $x_j^{(l)}$

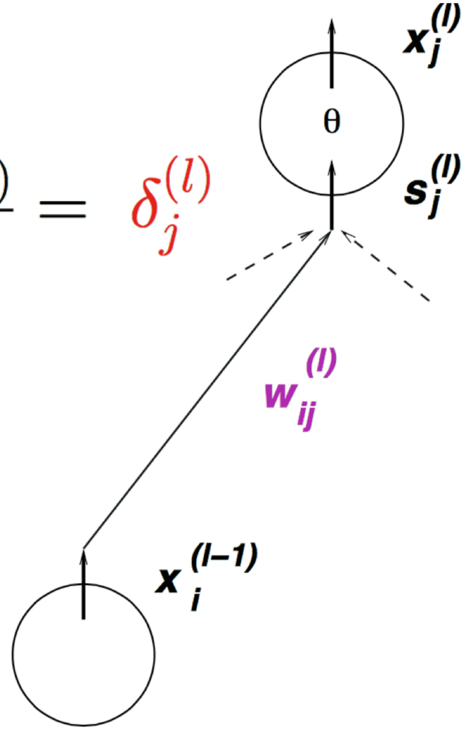
  Backward: Compute all $\delta_j^{(l)}$

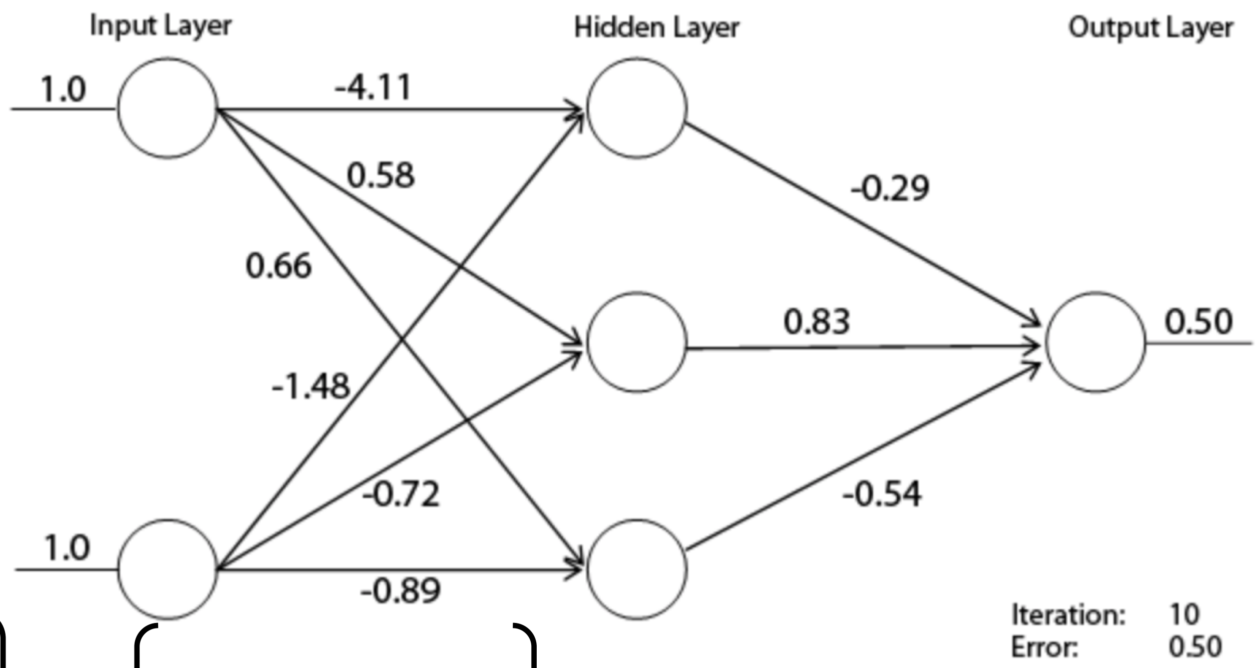
  Update the weights: $w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta\, x_i^{(l-1)} \delta_j^{(l)}$

  Iterate to the next step until it is time to stop

Return the final weights $w_{ij}^{(l)}$

$$\frac{\partial\, \mathrm{e}(\mathbf{w})}{\partial\, s_j^{(l)}} = \delta_j^{(l)}$$



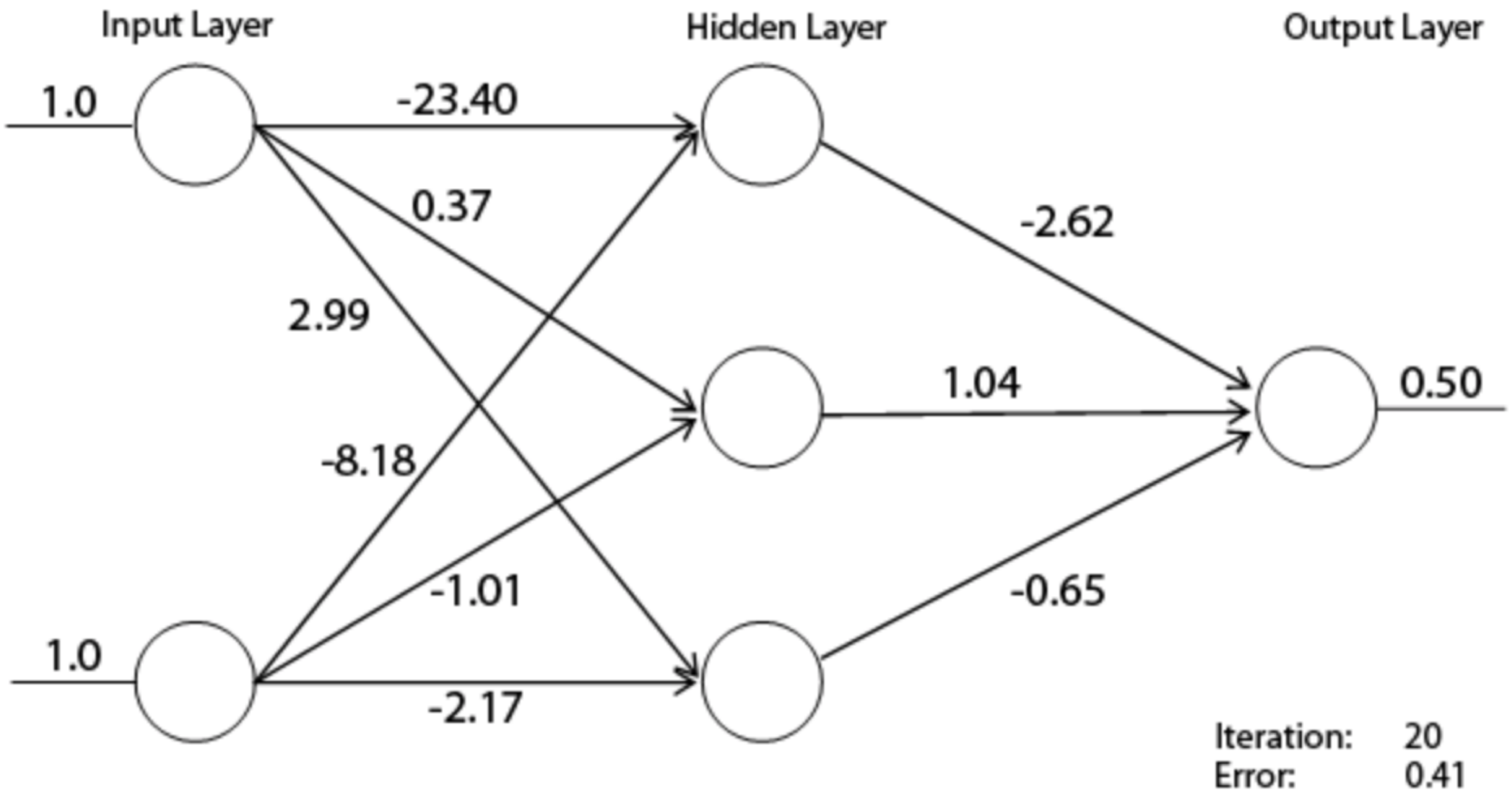


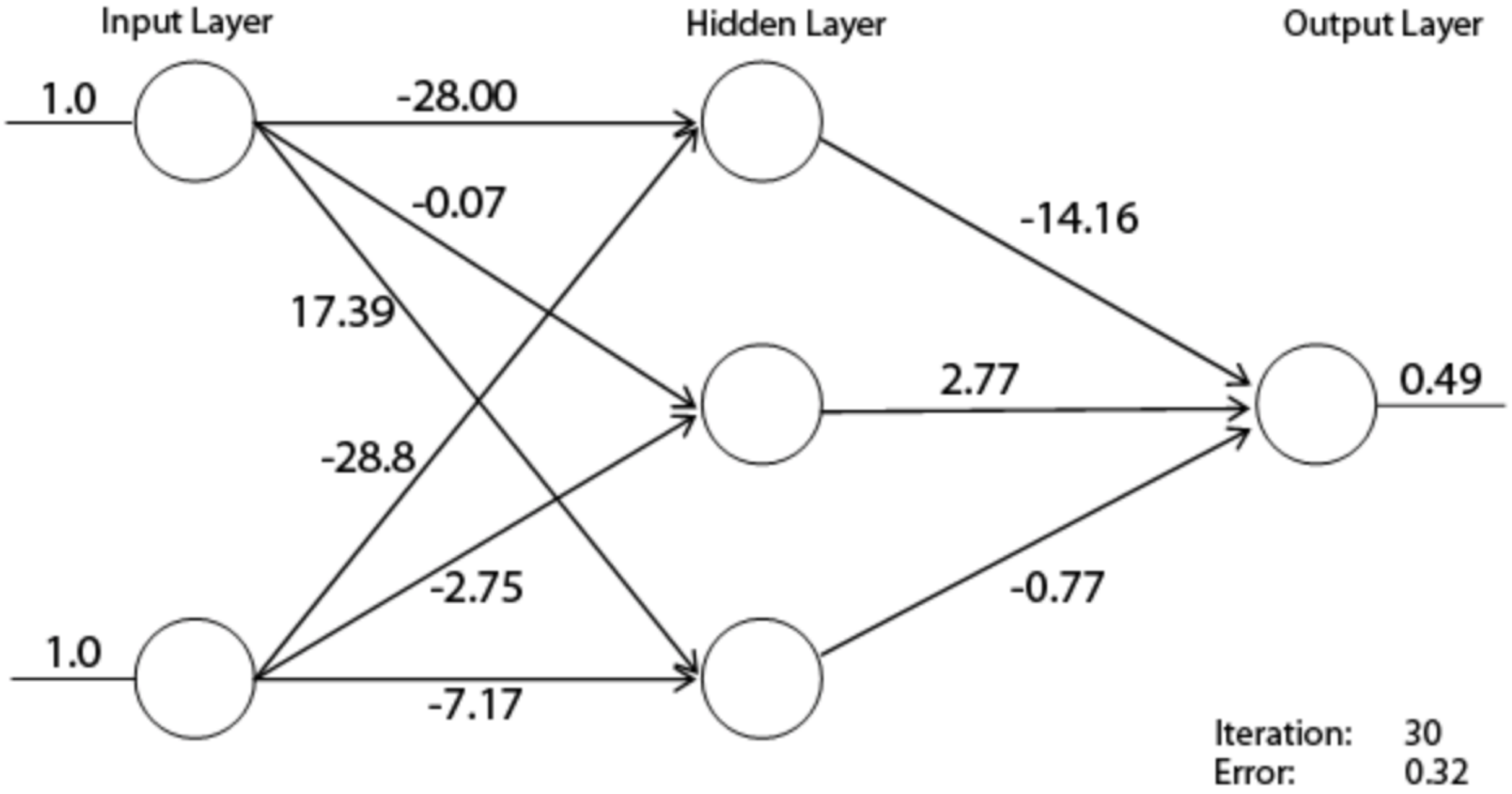
$$\begin{pmatrix} 1.0. & 1.0 \end{pmatrix} \begin{pmatrix} w11 & w12 & w13 \\ w21 & w22 & w23 \end{pmatrix} = \begin{pmatrix} y1 & y2 & y3ç \end{pmatrix}$$

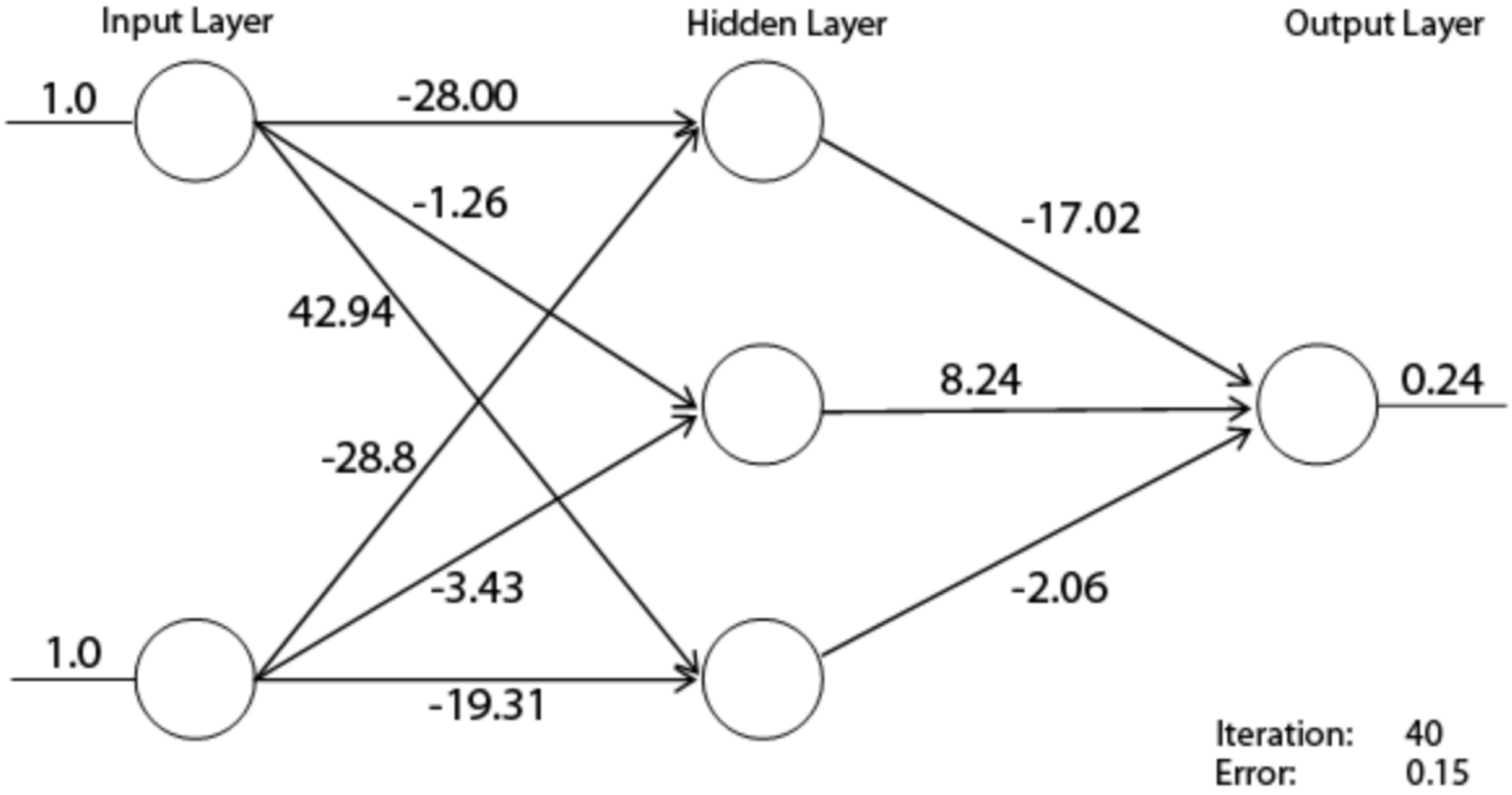
$$\begin{pmatrix} y1 & y2 & y3ç \end{pmatrix} \begin{pmatrix} w1 \\ W2 \\ w3 \end{pmatrix} = \begin{pmatrix} 0.50 \end{pmatrix}$$

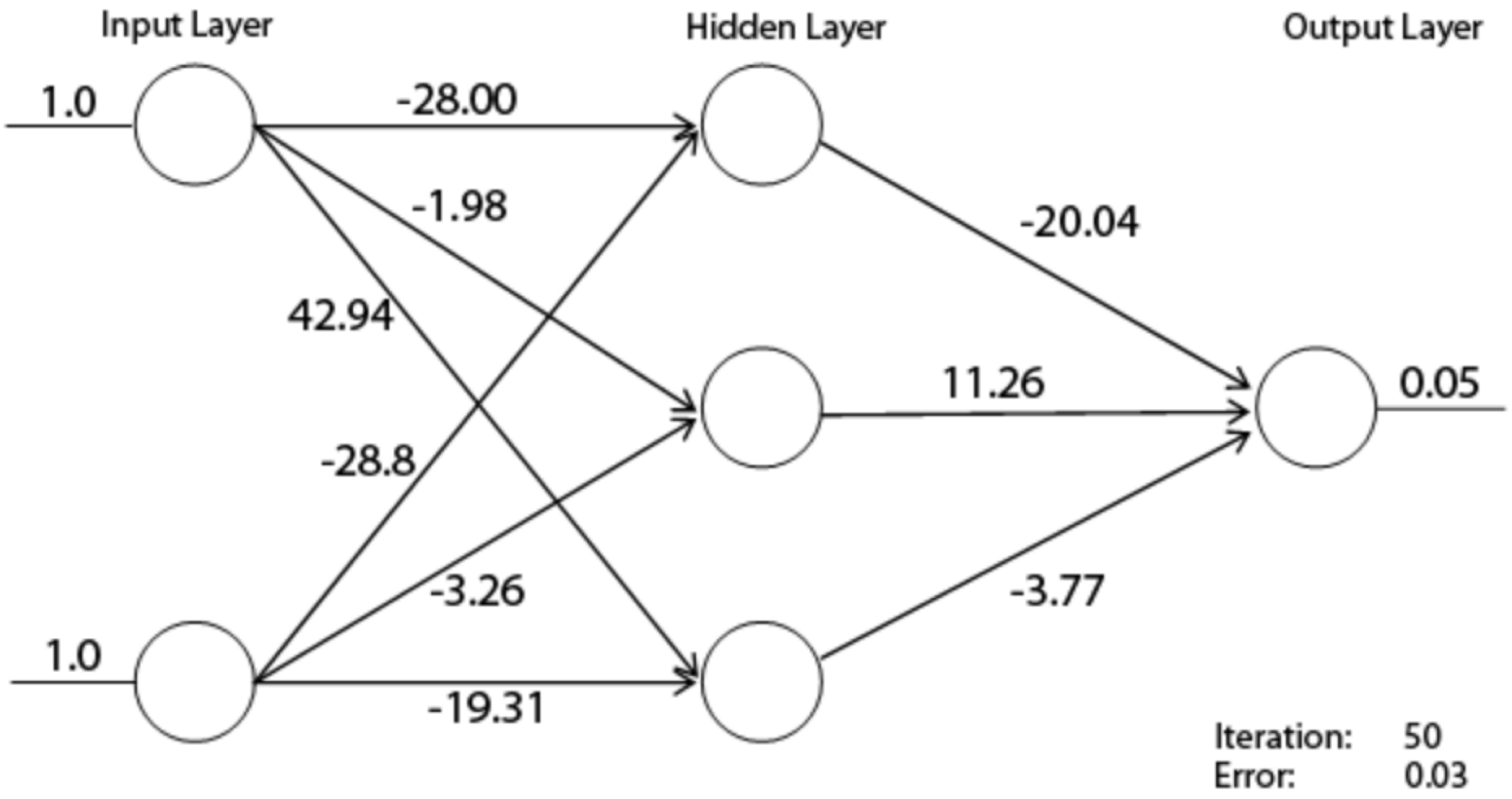
Y = 0.0
Y-hat = 0.749

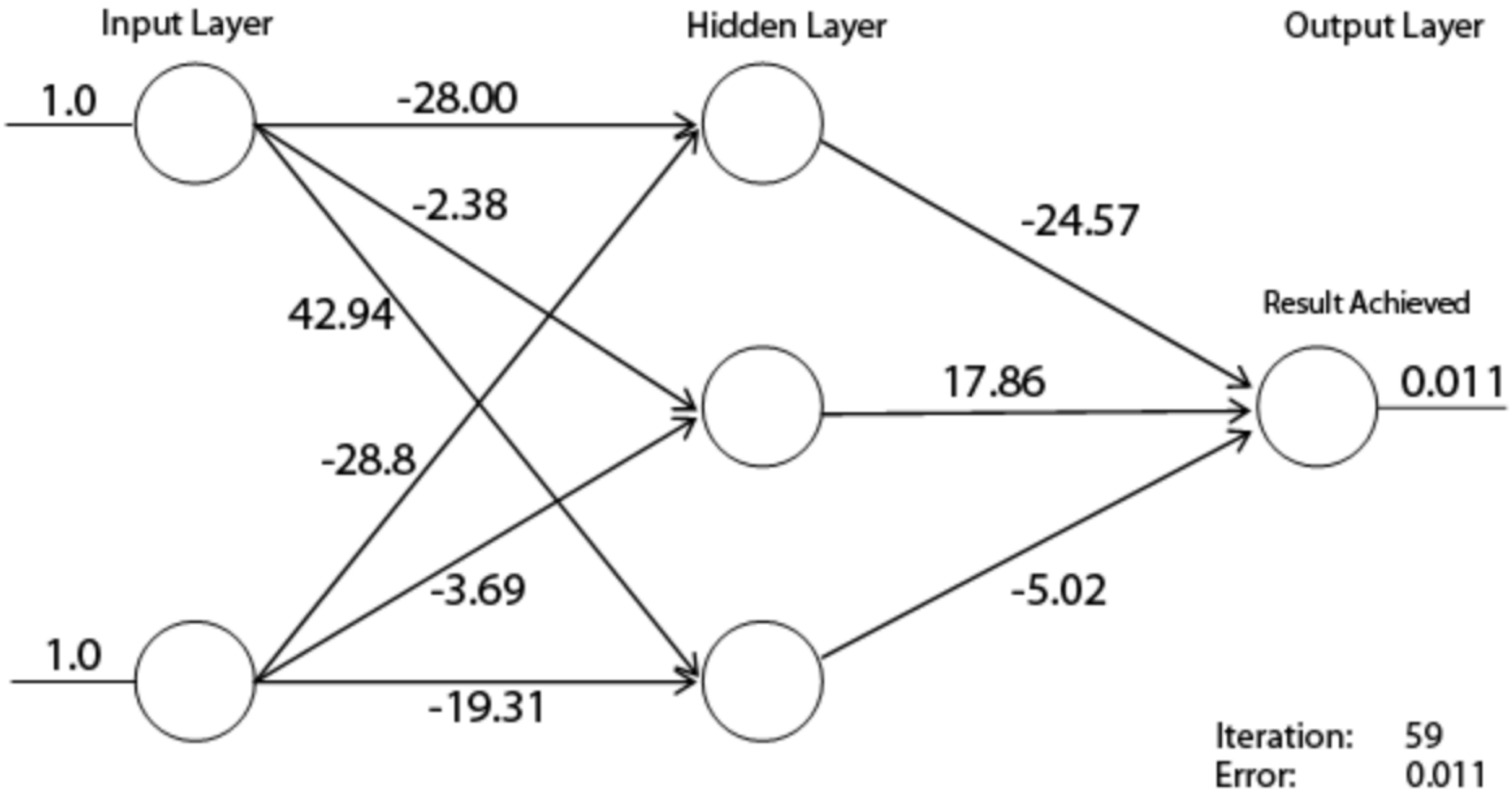
Error = 0.50 – 0 = 0.50


Input Layer
Hidden Layer
Output Layer
1.0
-23.40
0.37
2.99
-8.18
-1.01
-2.17
1.0
-2.62
1.04
-0.65
0.50
Iteration: 20
Error: 0.41


Input Layer
Hidden Layer
Output Layer
1.0
1.0
-28.00
-0.07
17.39
-28.8
-2.75
-7.17
-14.16
2.77
-0.77
0.49
Iteration: 30
Error: 0.32


Input Layer
Hidden Layer
Output Layer
1.0
-28.00
-1.26
-17.02
42.94
8.24
0.24
-28.8
-3.43
-2.06
1.0
-19.31
Iteration: 40
Error: 0.15


Input Layer
Hidden Layer
Output Layer
1.0
1.0
-28.00
-1.98
42.94
-28.8
-3.26
-19.31
-20.04
11.26
-3.77
0.05
Iteration: 50
Error: 0.03

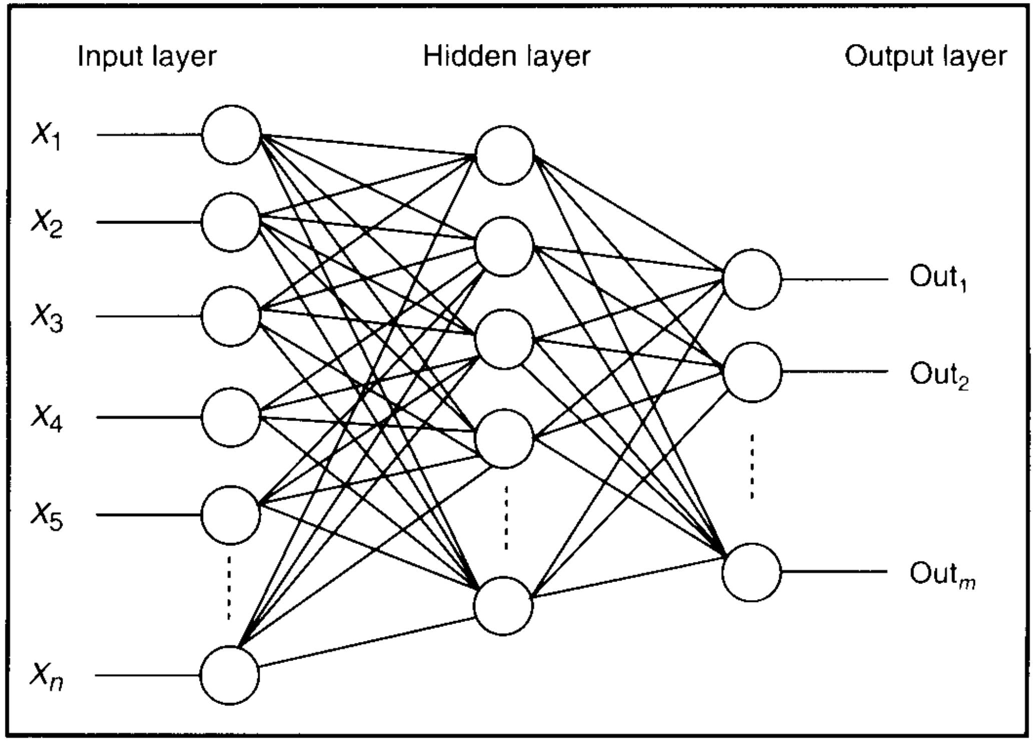

Input Layer
Hidden Layer
Output Layer
1.0
-28.00
-2.38
42.94
-24.57
Result Achieved
17.86
0.011
-28.8
-3.69
-5.02
1.0
-19.31
Iteration: 59
Error: 0.011

Learning Rate: Steps

The image part with relationship ID rId2 was not found in the file.

Learning Rate should increase with the slope

Example: Multi output units



0

1

…

9

$$h(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$h(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$h(x) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$