AI: Neural Networks

CPSC 501: Advanced Programming Techniques
Fall 2020

Jonathan Hudson, Ph.D
Instructor
Department of Computer Science
University of Calgary

Wednesday, August 12, 2020

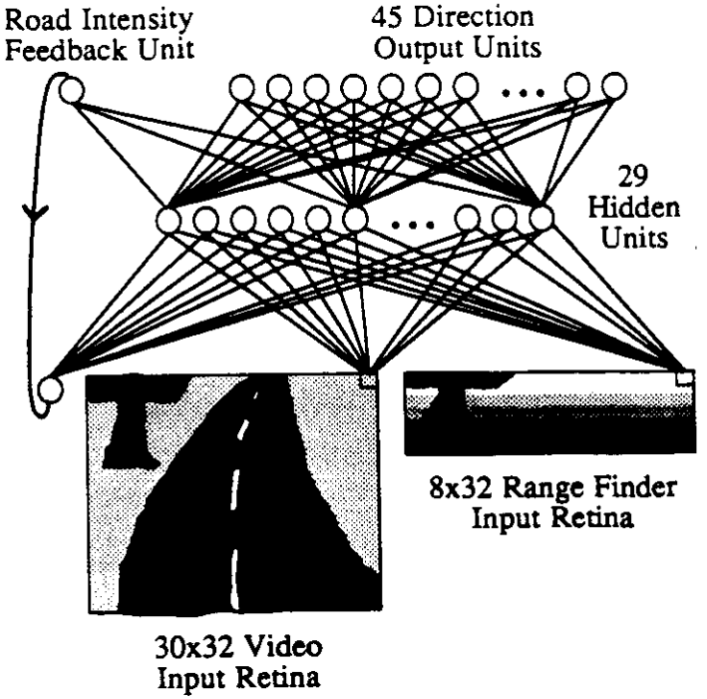
UNIVERSITY OF CALGARY

A short history of Neural Networks

- 1957: Perceptron (Frank Rosenblatt): one layer network neural network
- 1959: first neural network to solve a real world problem, i.e., eliminates echoes on phone lines (Widrow & Hoff)
- **First AI Winter**
- 1988: Backpropagation (Rumelhart, Hinton, Williams): learning a multi-layered network
- **Second AI Winter**

A short history of NNs

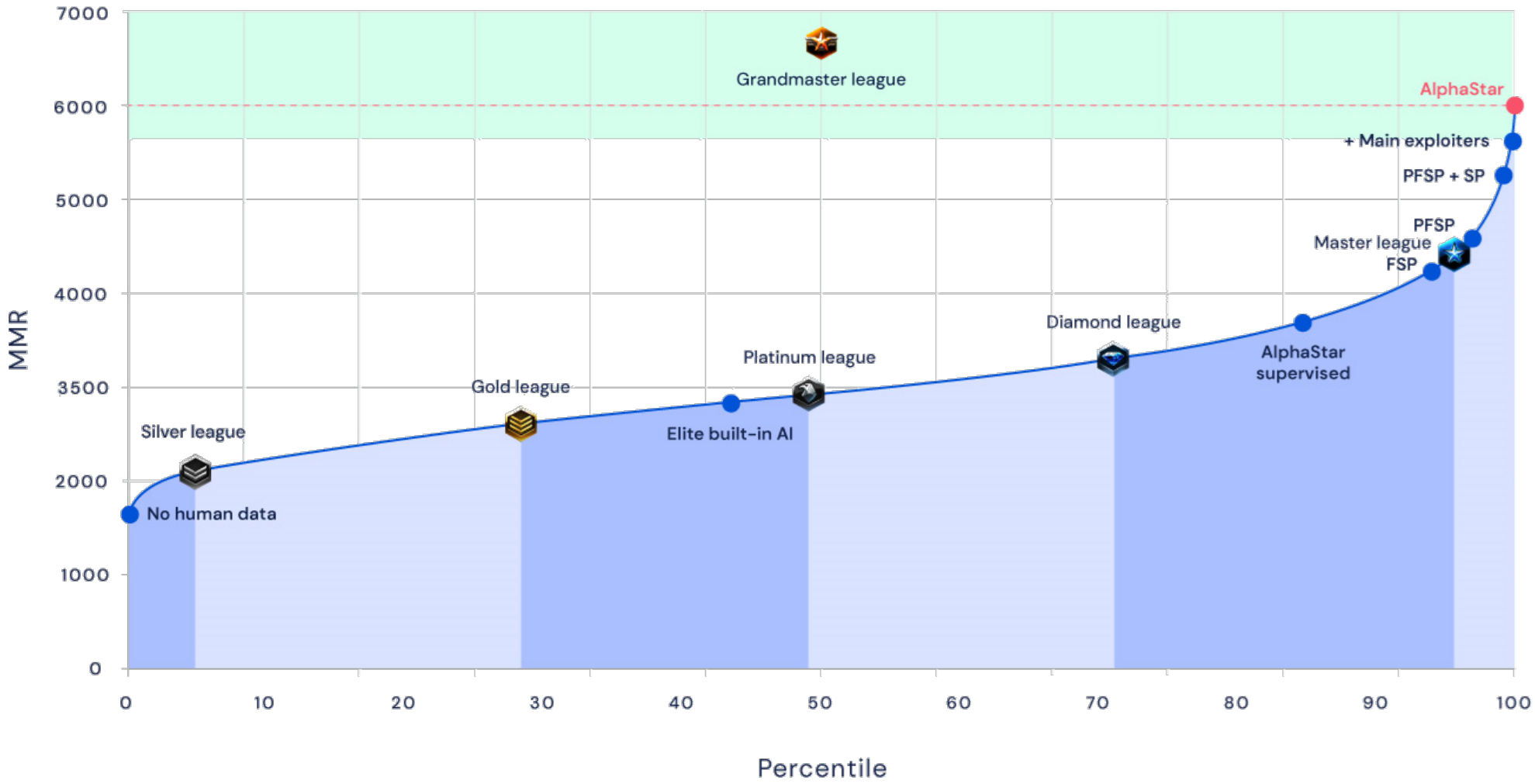
- 1989: ALVINN: autonomous driving car using NN (CMU)



A short history of NNs

- 1989: (LeCun) Successful application to recognize handwritten ZIP codes on mail using a “deep” network
- 2010s: near-human capabilities for image recognition, speech recognition, and language translation
- 2019 (AlphaStar) Google’s StarCraft 2 AI better than 99.8% of human players (GrandMaster level)

A short history of NNs



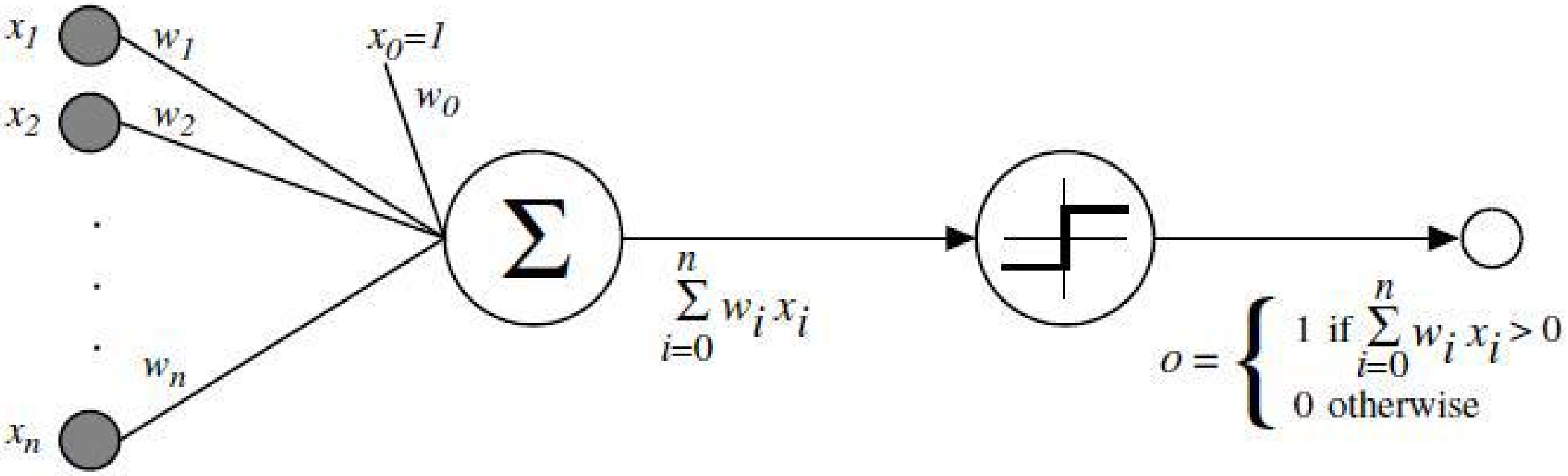
UNIVERSITY OF CALGARY

Perceptron

Not a Dr. Who Robot (or maybe it is?)

Perceptron

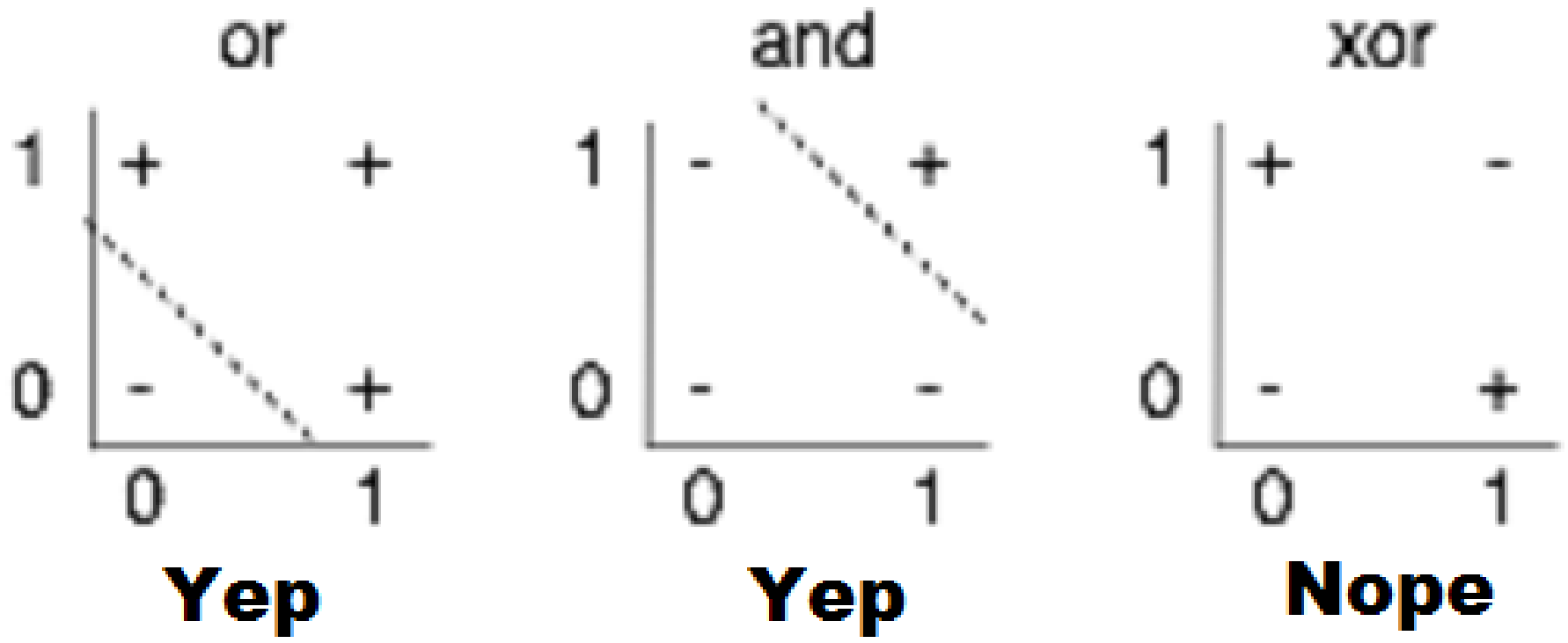
- Invented by Frank Rosenblatt (1957): simplified mathematical model of how the neurons in our brains operate



From: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

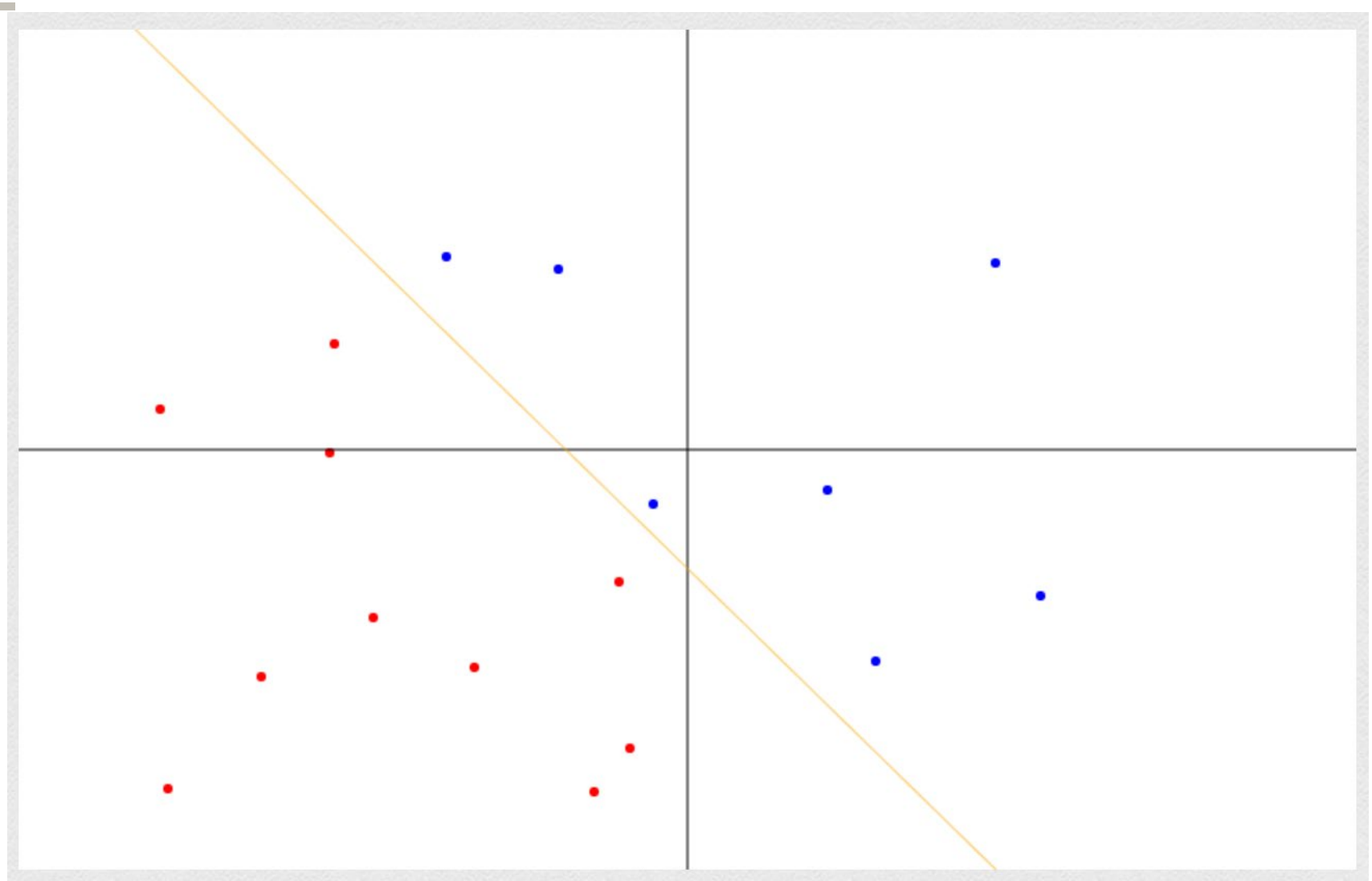
Perceptron

- Could implement AND, OR, but not XOR



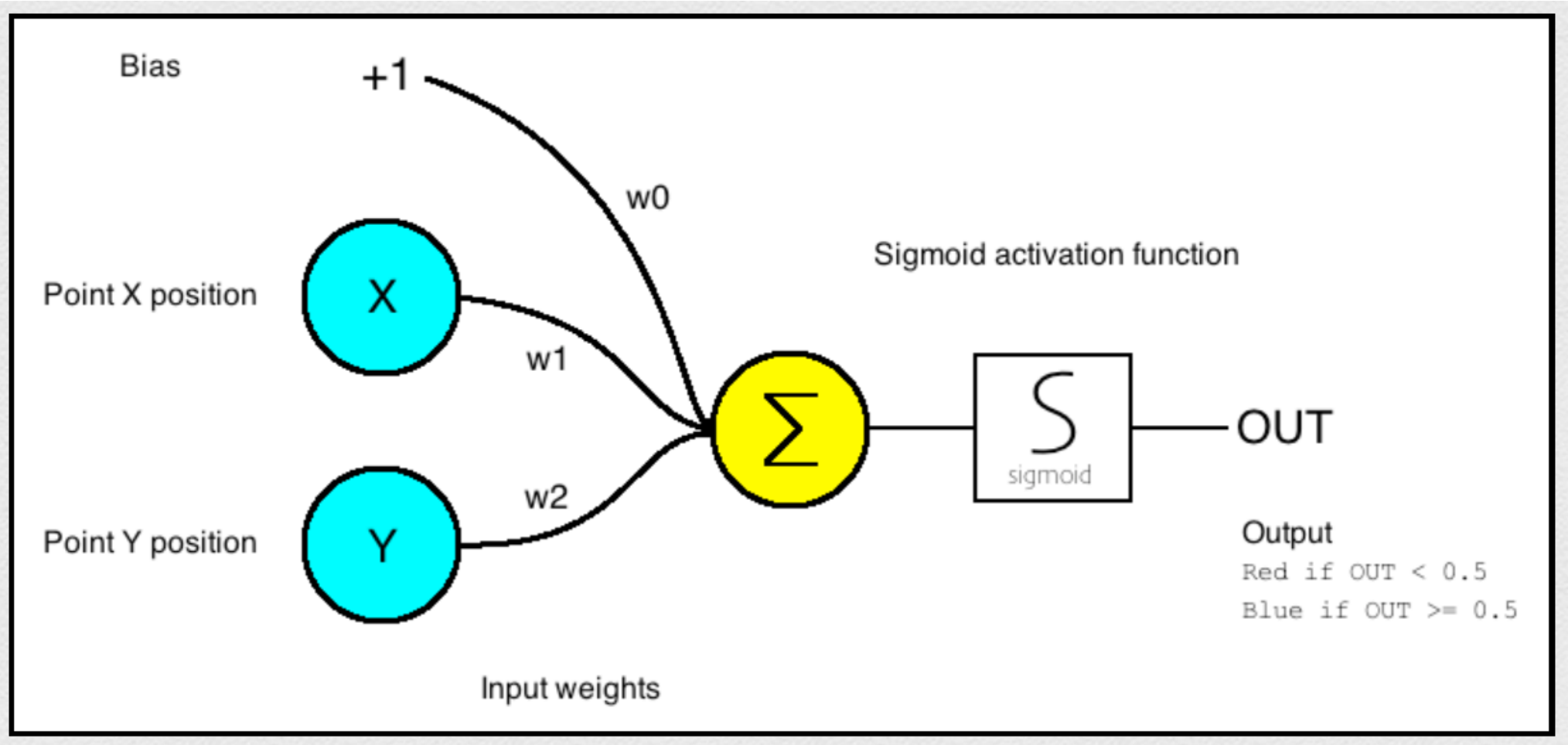
From: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

Perceptron



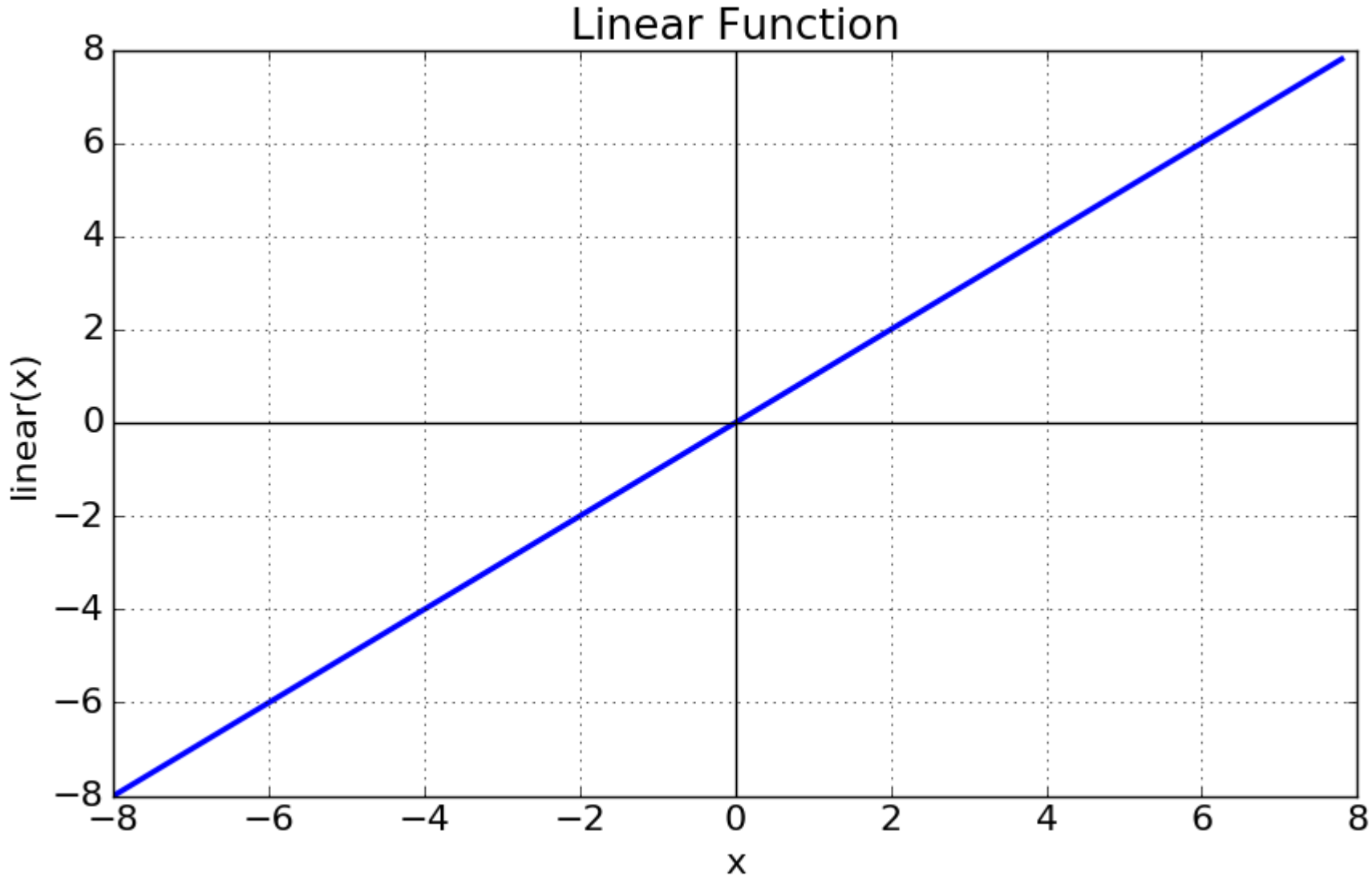
https://www.cs.utexas.edu/~teammco/misc/perceptron/

Perceptron

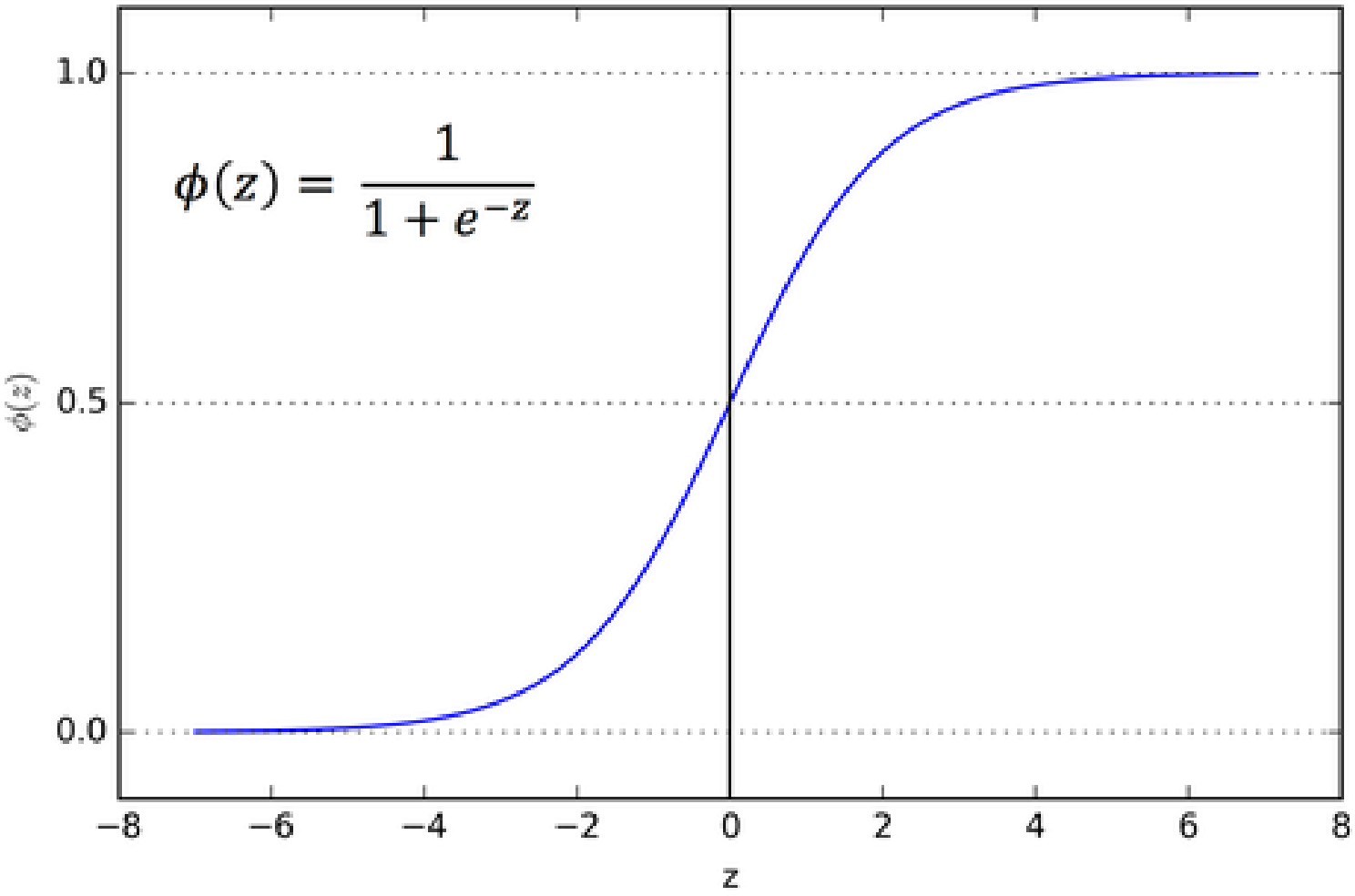


Activation Functions

Activation Functions (Identity)



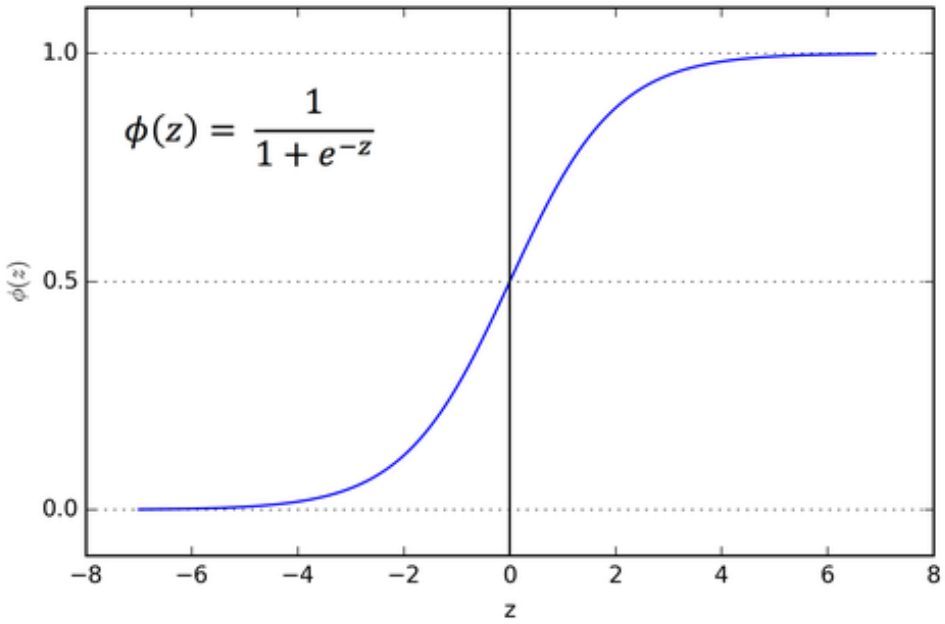
Sigmoid Activation Function



UNIVERSITY OF CALGARY

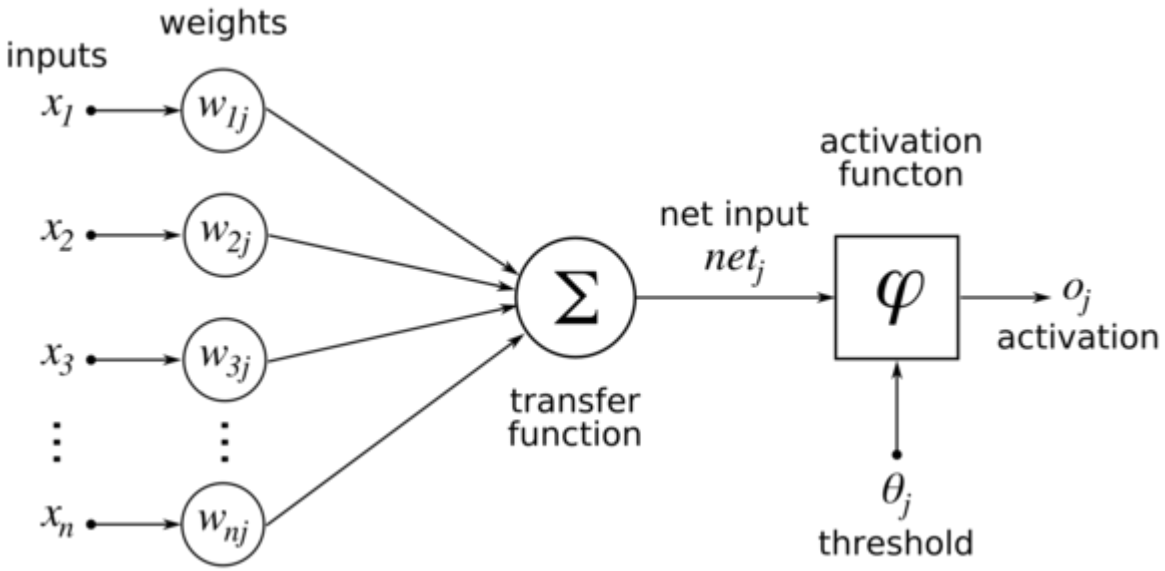
Sigmoid Activation Function

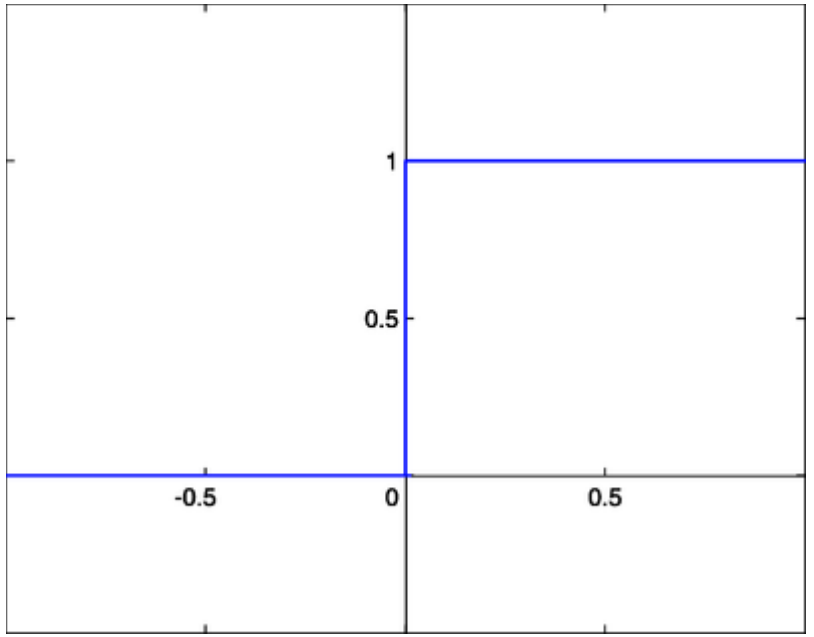
- The main reason why we use sigmoid function is because gradations exist between **(0 to 1)**
- Used for models where we have to **predict the probability** as an output.



Step (Sigmoid) Activation Function

- Replace a smooth function with a critical threshold point (if value exceeds then fully one answer)

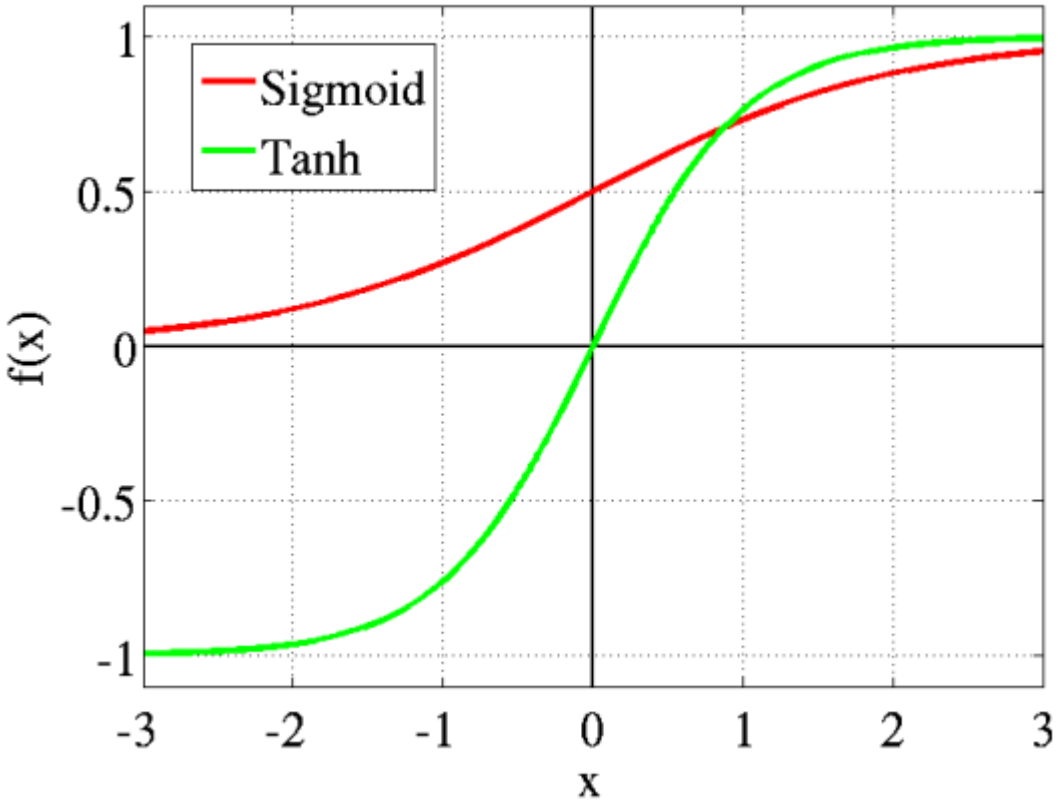




UNIVERSITY OF CALGARY

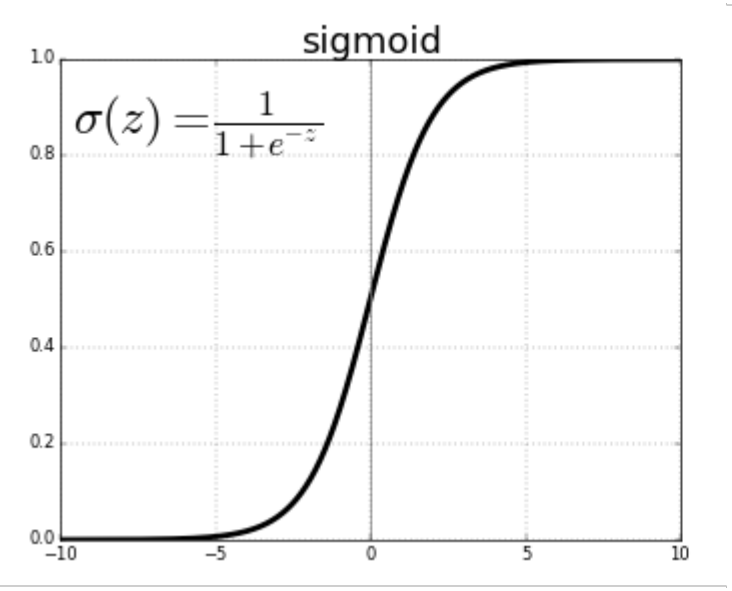
Tanh Activation Function

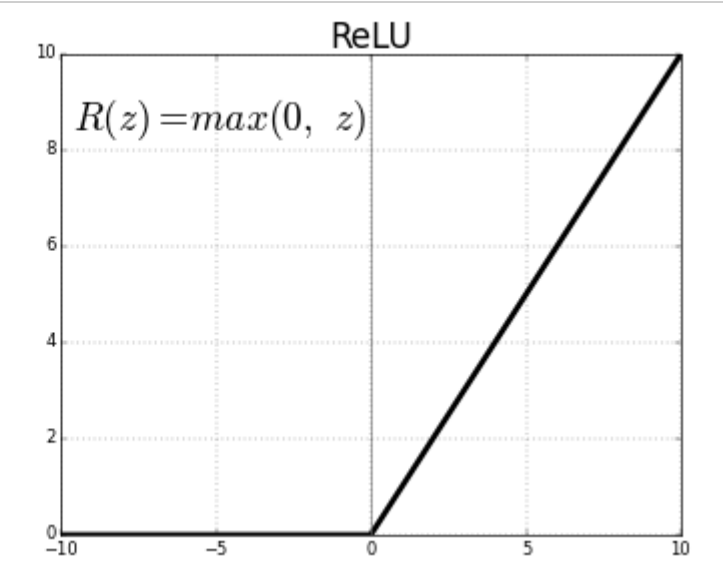
- The advantage is that the negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the tanh graph.
- The tanh function is mainly used classification between two classes.



ReLU (Rectified Linear Unit) Activation Function

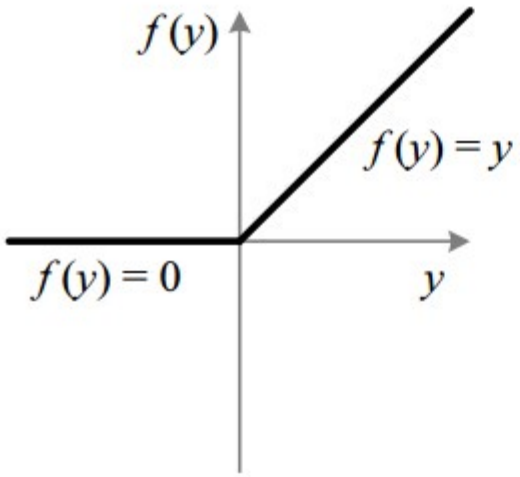
- The ReLU is the most used activation function in the world right now.
- Most used, due to Convolution Neural Network popularity
- But the issue is that all the negative values become zero immediately which decreases the ability of the model to fit or train from the data properly.

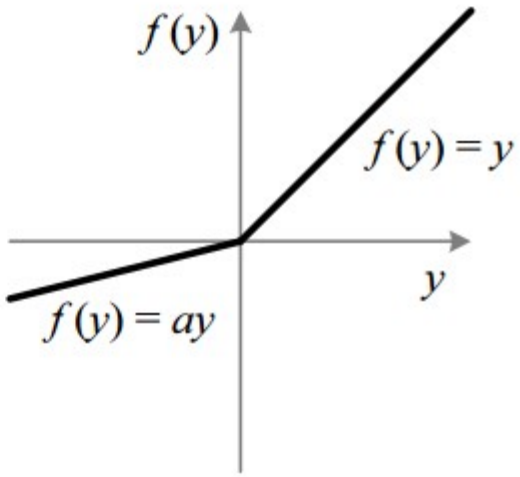


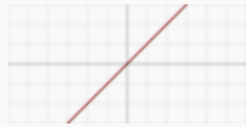
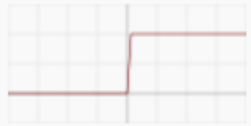
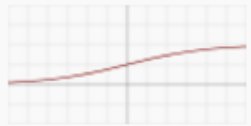
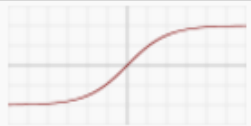
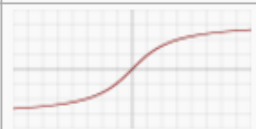
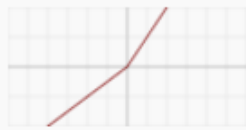
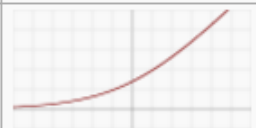


Leaky ReLU Activation Function

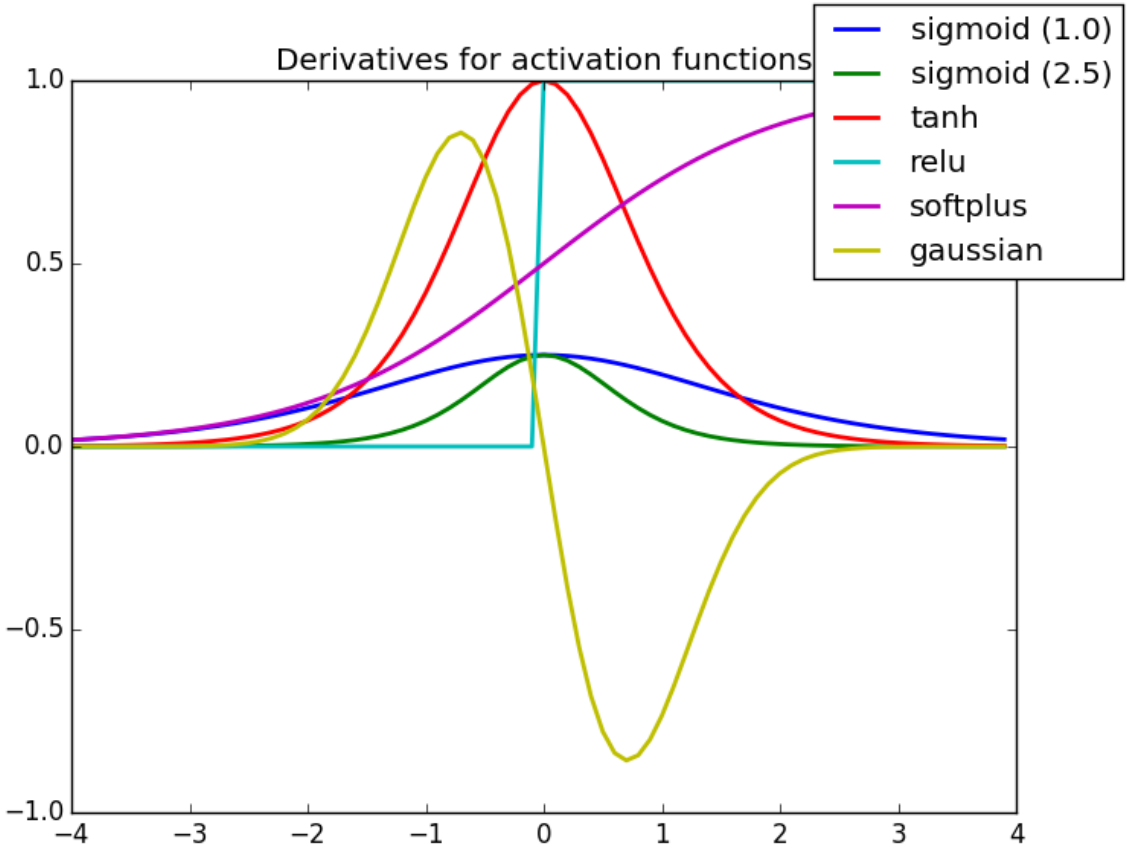
- It is an attempt to solve the dying ReLU problem
- The leak helps to increase the range of the ReLU function. Usually, the value of **a** is 0.01 or so.





Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for} \quad x \neq 0 \\ ? & \text{for} \quad x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$
Parameteric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

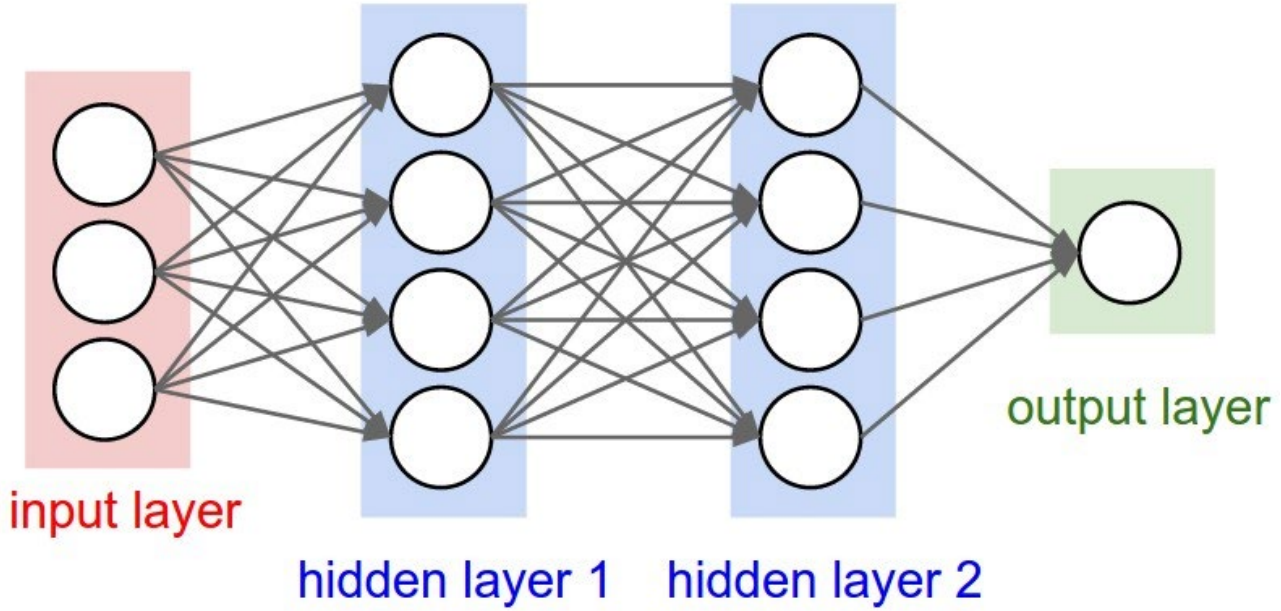
Derivatives of Activation Functions



Layers

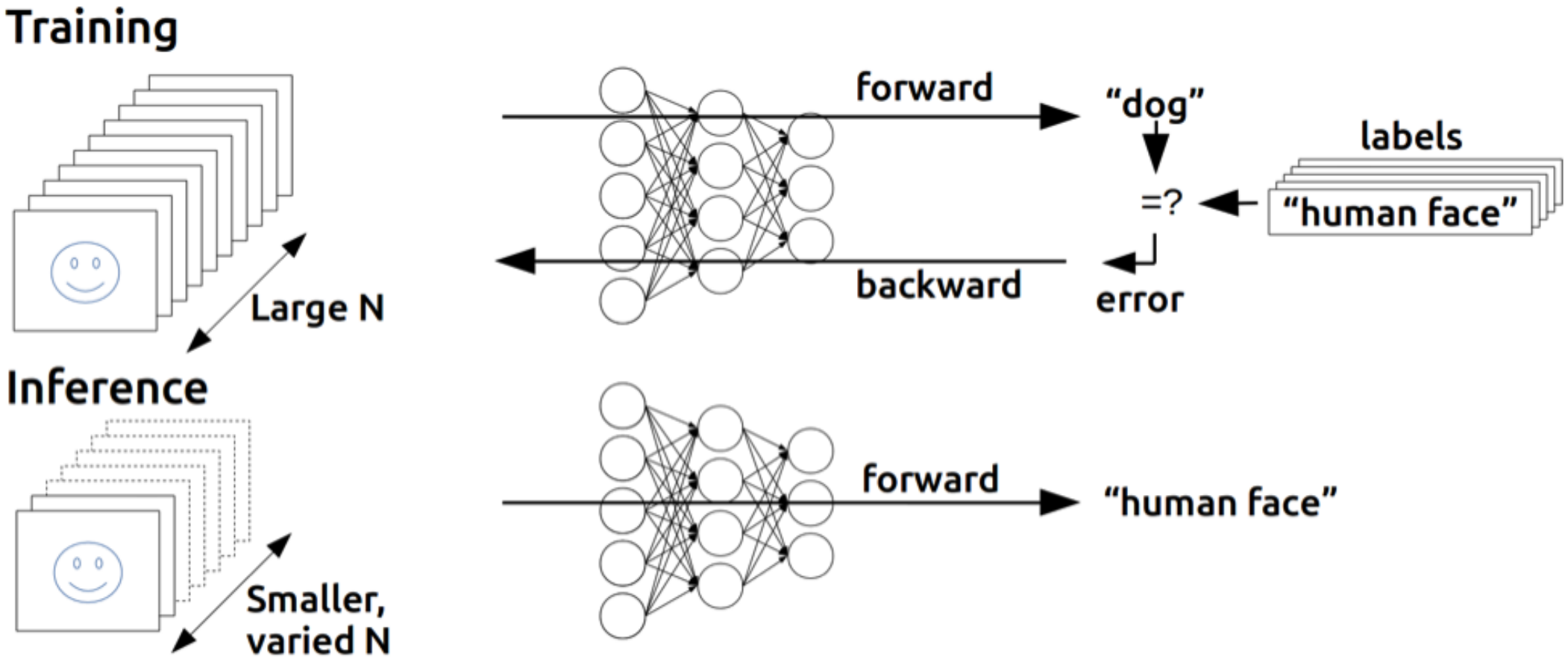
Hidden layers

- Hidden layers can find **features** within the data and allow following layers to operate on those features
  - Can implement XOR



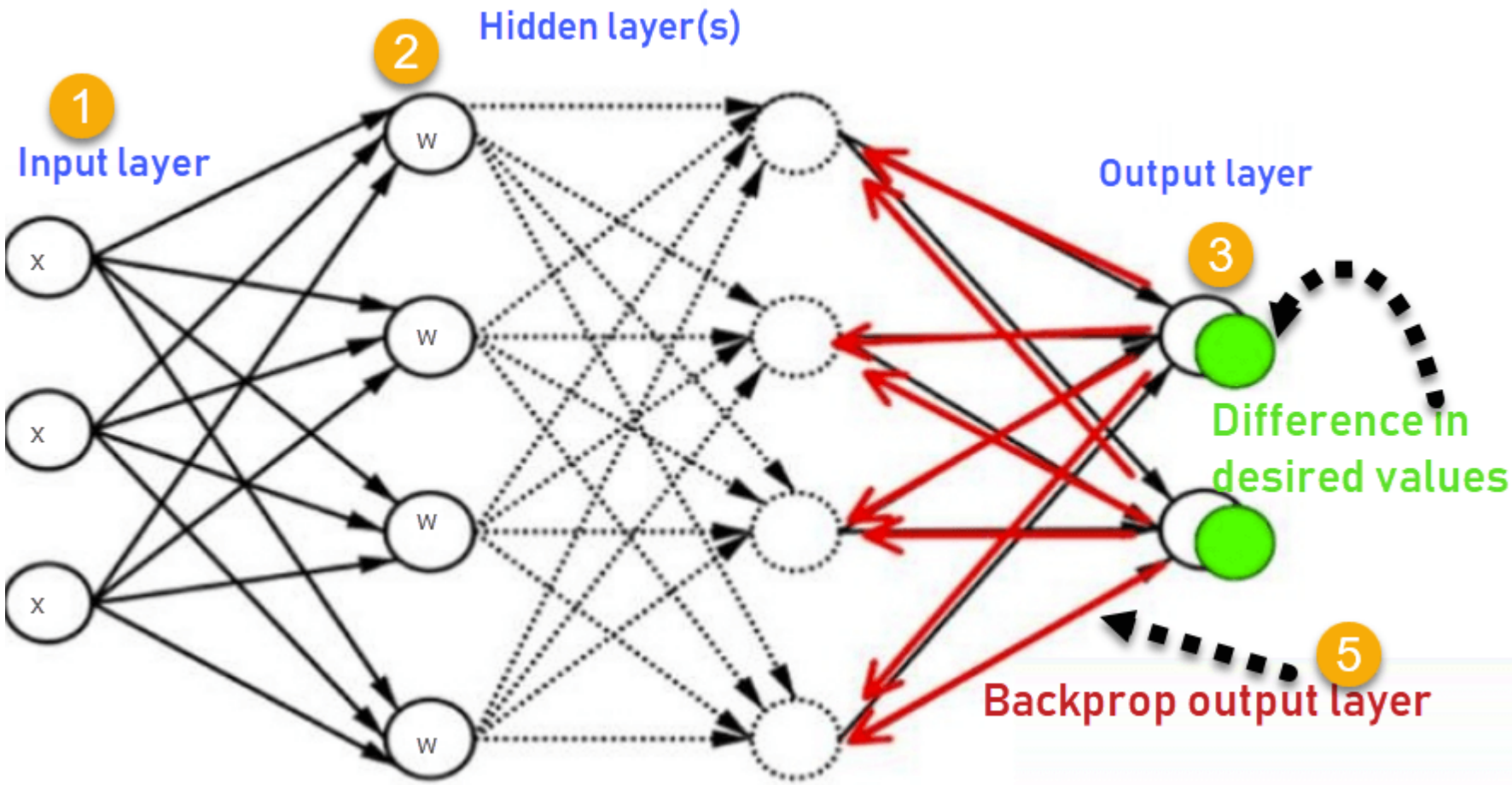
From: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

Learning: Backpropagation

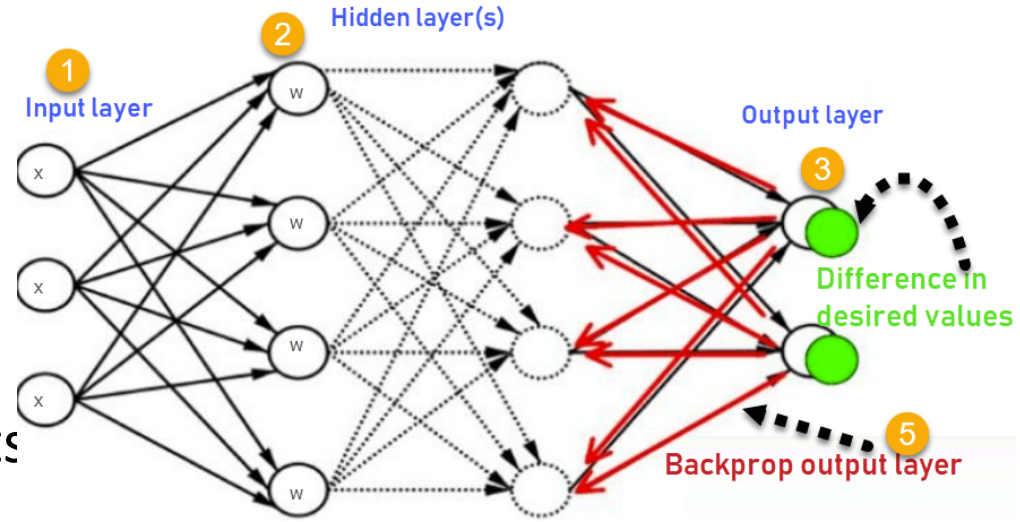


From: http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

Back-Propagation



Back-Propagation



- Inputs X, arrive through the preconnected path
- Input is modeled using real weights W. The weights selected.
- Calculate the output for every neuron from the input layer, to the hidden layers, to the output layer.
- Calculate the error in the outputs
- ErrorB= Actual Output – Desired Output
- Travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased.

Context (circa 2015)

- Deep learning already claiming big successes

Team	Year	Place	Error (top-5)
XRCE (pre-neural-net explosion)	2011	1st	25.8%
Supervision (AlexNet)	2012	1st	16.4%
Clarifai	2013	1st	11.7%
GoogLeNet (Inception)	2014	1st	6.66%
Andrej Karpathy (human)	2014	N/A	5.1%
BN-Inception (Arxiv)	2015	N/A	4.9%
Inception-v3 (Arxiv)	2015	N/A	3.46%

Imagenet challenge classification task

From: http://www.wsdm-conference.org/2016/slides/WSDM2016-Jeff-Dean.pdfz

Lesson: ImageNet

ImageNet

- Not designed for people
- Recently went viral
- Sept 23, 2019
- “ImageNet will remove 600,000 images of people stored on its database after an art project exposed racial bias in the program’s artificial intelligence system.”

ImageNet

- First presented as a research poster in 2009
- Scraped a collection of many millions of images from the internet
- Trained through images categorized by Amazon Mechanical Turk workers
- Crowdsourcing platform through which people can earn money performing small tasks
- Sorted an average of 50 images per minute into thousands of categories
- In 2012, a team from the University of Toronto used a Convolutional Neural Network to handily win the top prize
- Final year 2017, and accuracy in classifying objects in the limited subset had risen from 71.8% to 97.3%. That did not include “Person” category

ImageNet

- AI researcher Kate Crawford and artist Trevor Paglen
  - Training Humans — an exhibition that at the Prada Foundation in Milan
  - Part of their experiment also lives online at ImageNet Roulette, a website where users can upload their own photographs to see how the database might categorize them.
  - https://www.excavating.ai/
- Example of the complexities and dangers of human classification
- The sliding spectrum between supposedly unproblematic labels like “trumpeter” or “tennis player” to concepts like “spastic,” “mulatto,” or “redneck.”
- ImageNet is an object lesson in what happens when people are categorized like objects.

Onward to …
TensorFlow.

Jonathan Hudson
jwhudson@ucalgary.ca
https://pages.cpsc.ucalgary.ca/~hudsonj/

UNIVERSITY OF CALGARY