

Intro to Deep Learning

Geena Kim

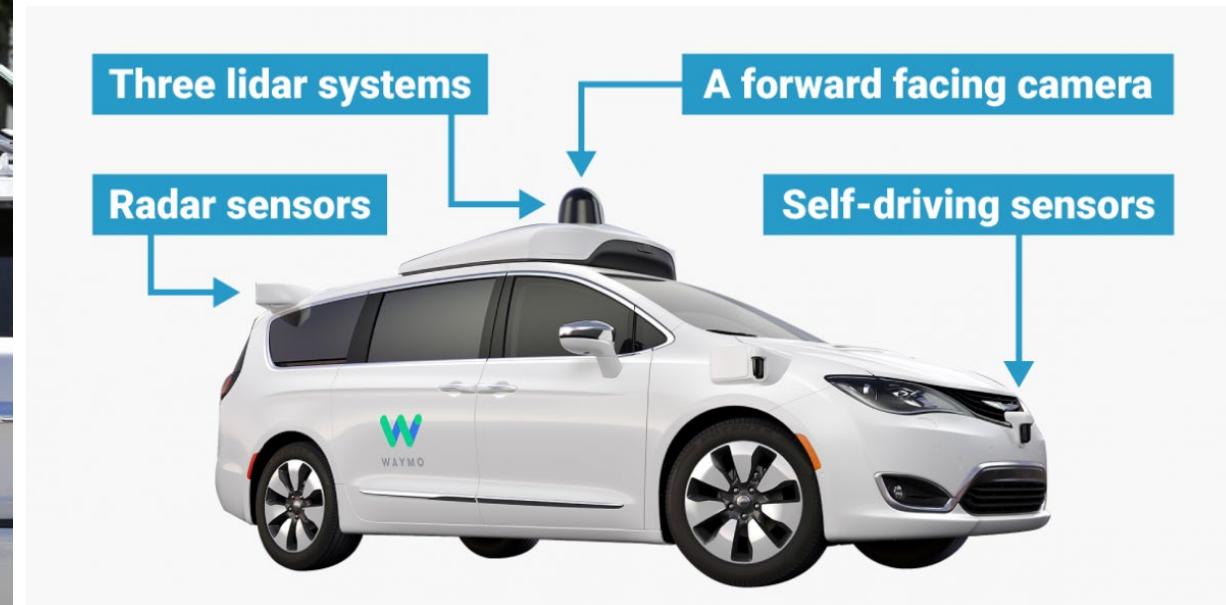
*Some of the slide/diagram adopted from CMU deep learning course



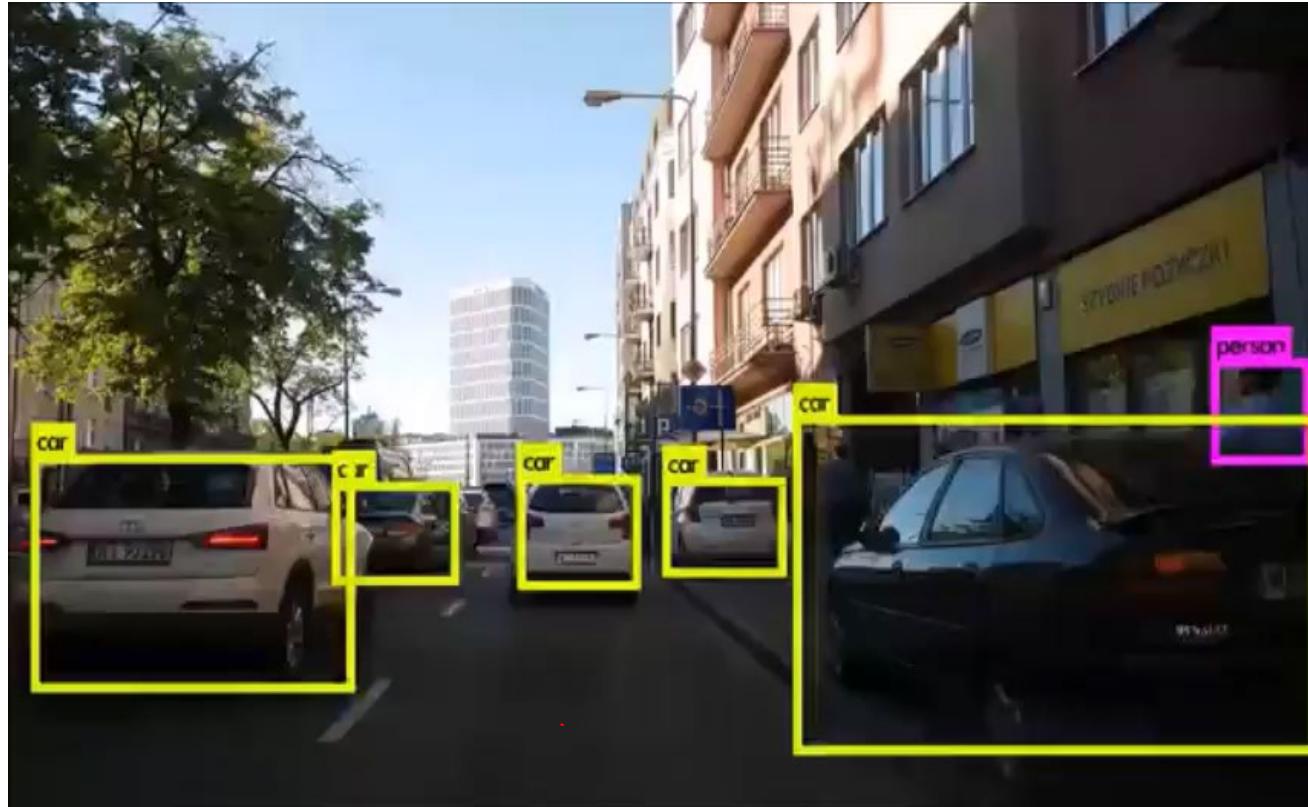
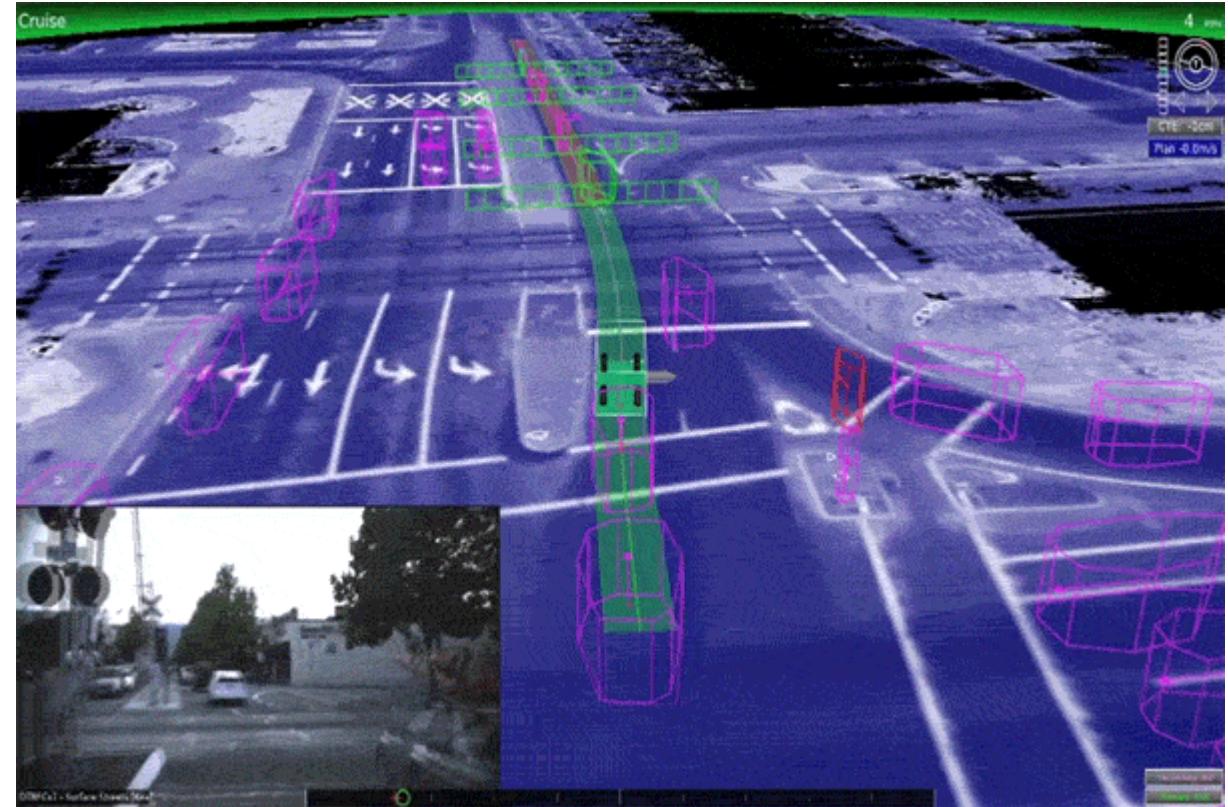
Why Deep Learning?



Self-driving Cars



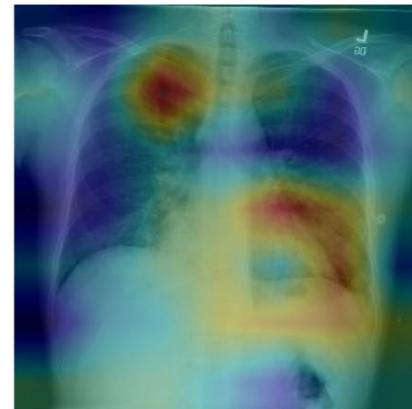
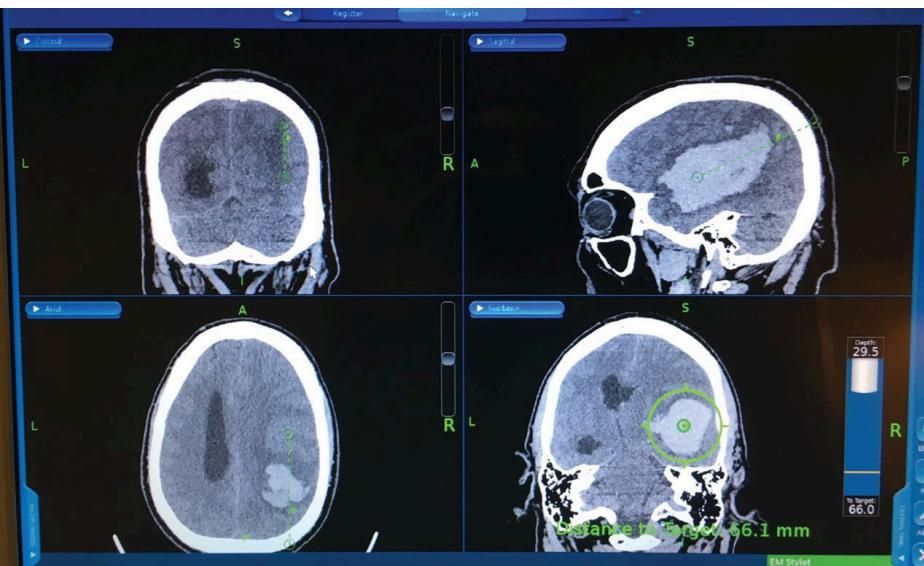
Why Deep Learning?



<https://www.youtube.com/watch?v=EhcpGpFHCrw>

Why Deep Learning?

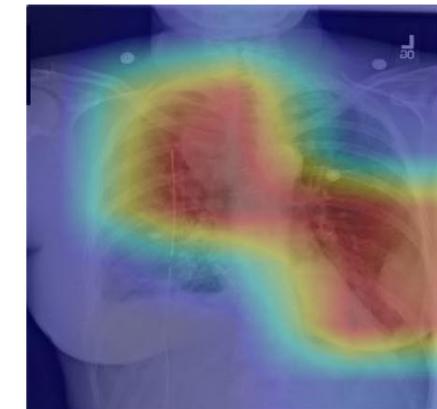
Medical Imaging, AI in Medicine



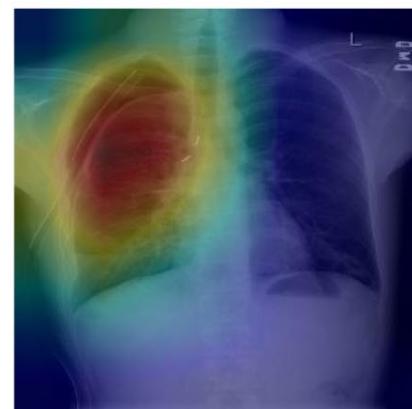
(a) Patient with multifocal community acquired pneumonia. The model correctly detects the airspace disease in the left lower and right upper lobes to arrive at the pneumonia diagnosis.



(b) Patient with a left lung nodule. The model identifies the left lower lobe lung nodule and correctly classifies the pathology.



(c) Patient with primary lung malignancy and two large masses, one in the left lower lobe and one in the right upper lobe adjacent to the mediastinum. The model correctly identifies both masses in the X-ray.



(d) Patient with a right-sided pneumothorax and chest tube. The model detects the abnormal lung to correctly predict the presence of pneumothorax (collapsed lung).



(e) Patient with a large right pleural effusion (fluid in the pleural space). The model correctly labels the effusion and focuses on the right lower chest.



(f) Patient with congestive heart failure and cardiomegaly (enlarged heart). The model correctly identifies the enlarged cardiac silhouette.
P. Rajpurkar et al., arXiv:1711.05225v3

Why Deep Learning?



News Startups Mobile Gadgets Enterprise Social Europe

Trending Amazon Tesla Microsoft

eBay

shopping

Search

eCommerce

eCommerce

Popular Posts

***Shopping,
e-commerce***

eBay launches visual search tools that let you shop using photos from your phone or web

Posted Oct 26, 2017 by **Sarah Perez** (@sarahintampa)



Next Story



Crunchbase

eBay

FOUNDED
1995

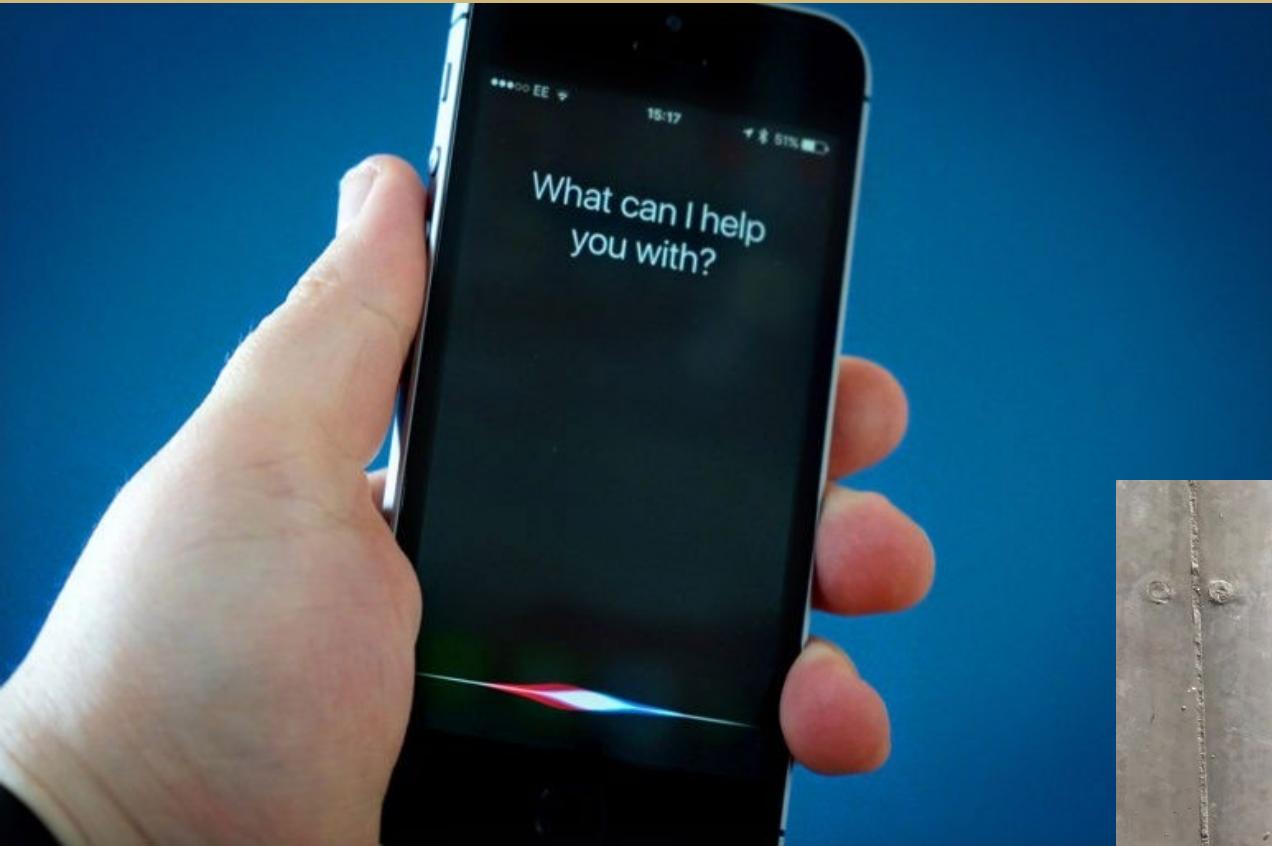
OVERVIEW

eBay is an online marketplace. The platform connects millions of buyers with sellers globally utilizing PayPal to ensure secure transactions. eBay products can be sold either via a silent auction in which users are able to input the maximum price they are willing to pay and for which the site will automatically increase bids as necessary up to that maximum, or via the Buy It

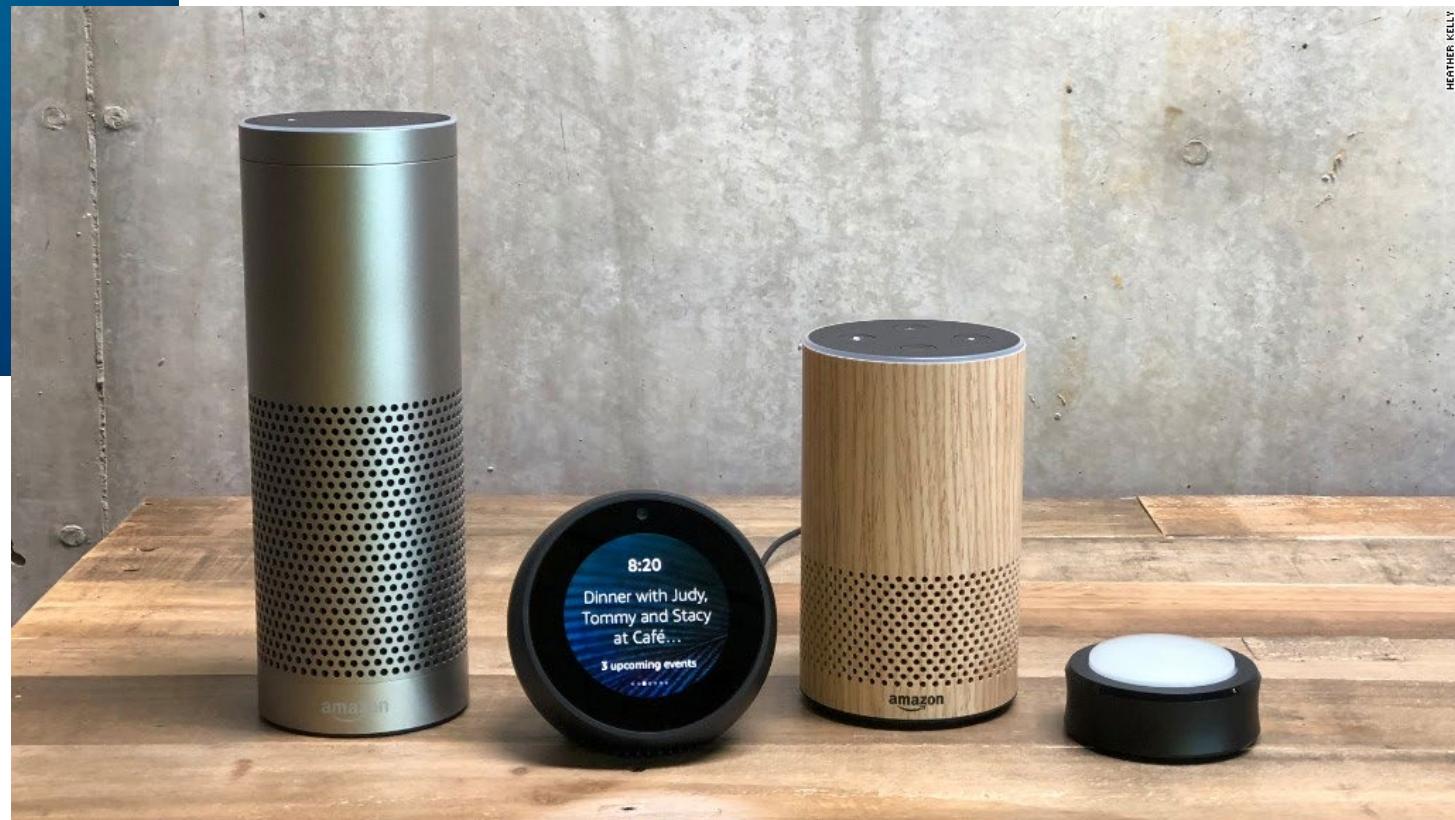
Computer vision in Robotics



Why Deep Learning?



*Voice Recognition,
Smart Devices*

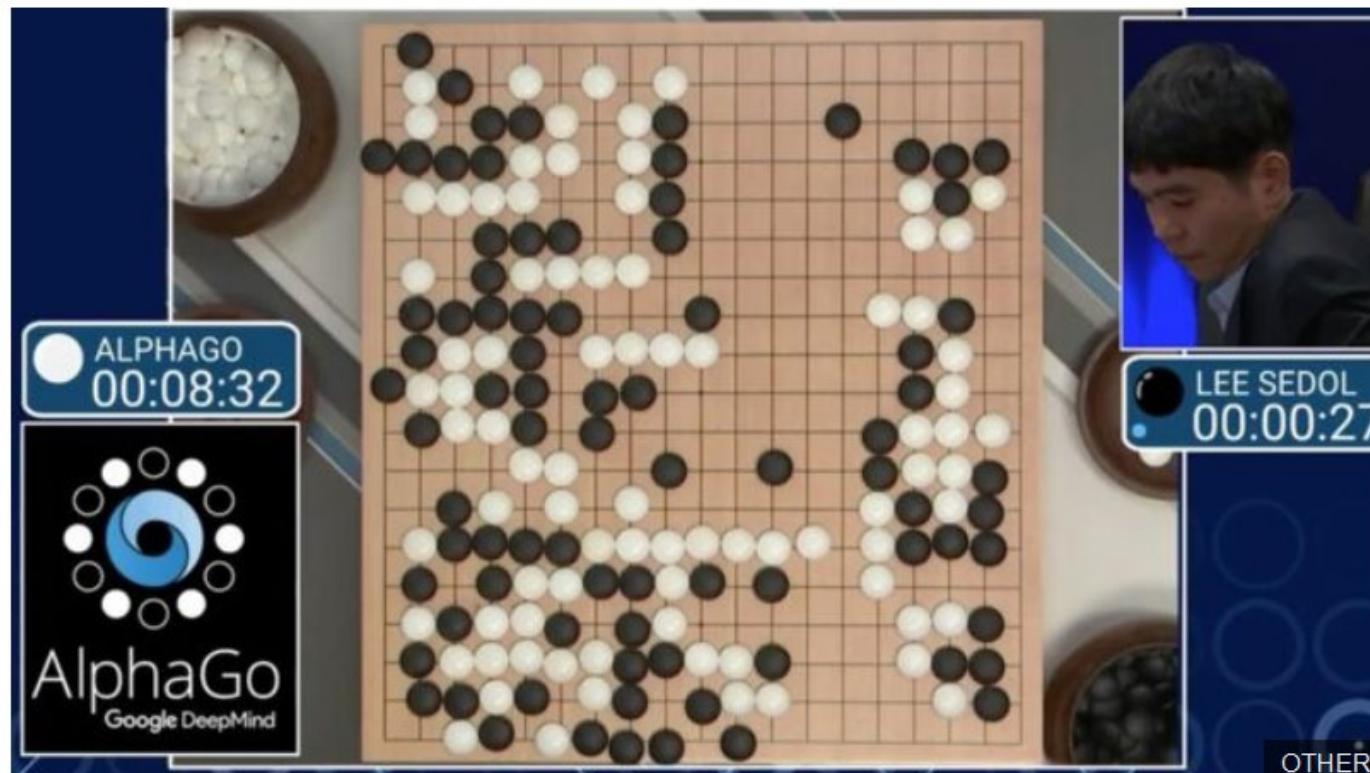


Deep learning in game playing

Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol

⌚ 12 March 2016

f Share



A computer program has beaten a master Go player 3-0 in a best-of-five competition, in what is seen as a landmark moment for artificial intelligence.

Deep learning face generation



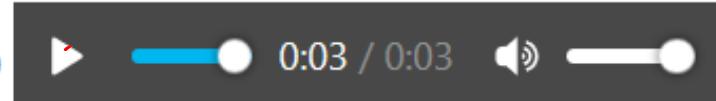
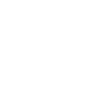
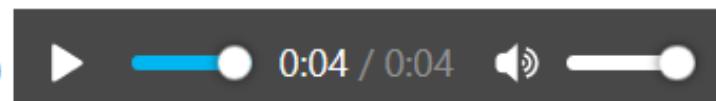
A few sample faces — all completely fake — created by ThisPersonDoesNotExist.com

Deep learning sound generation

The embedding can also transfer fine time-aligned prosody from one phrase to a slightly different phrase, though this technique works best when the reference and target phrases are similar in length and structure.

Reference Text: For the first time in her life she had been danced tired.

Synthesized Text: For the last time in his life he had been handily embarrassed.

Reference prosody (American)		
Synthesized without prosody embedding (American)		
Synthesized with prosody embedding (American)		

Excitingly, we observe prosody transfer even when the reference audio comes from a speaker whose voice is not in Tacotron's training data.

Deep learning explaining pictures

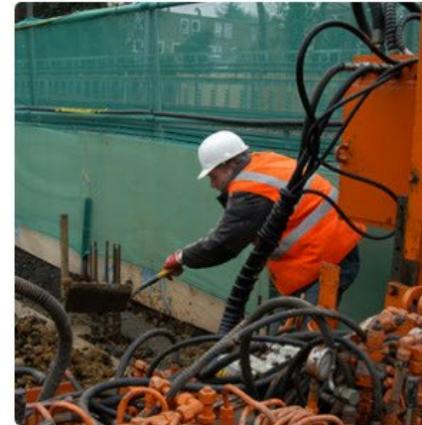
Multimodal Recurrent Neural Network

<https://cs.stanford.edu/people/karpathy/deepimagesent/>

Our Multimodal Recurrent Neural Architecture generates sentence descriptions from images. Below are a few examples of generated sentences:



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."



→ "girl in pink dress is jumping in air."



"black and white dog jumps over bar."



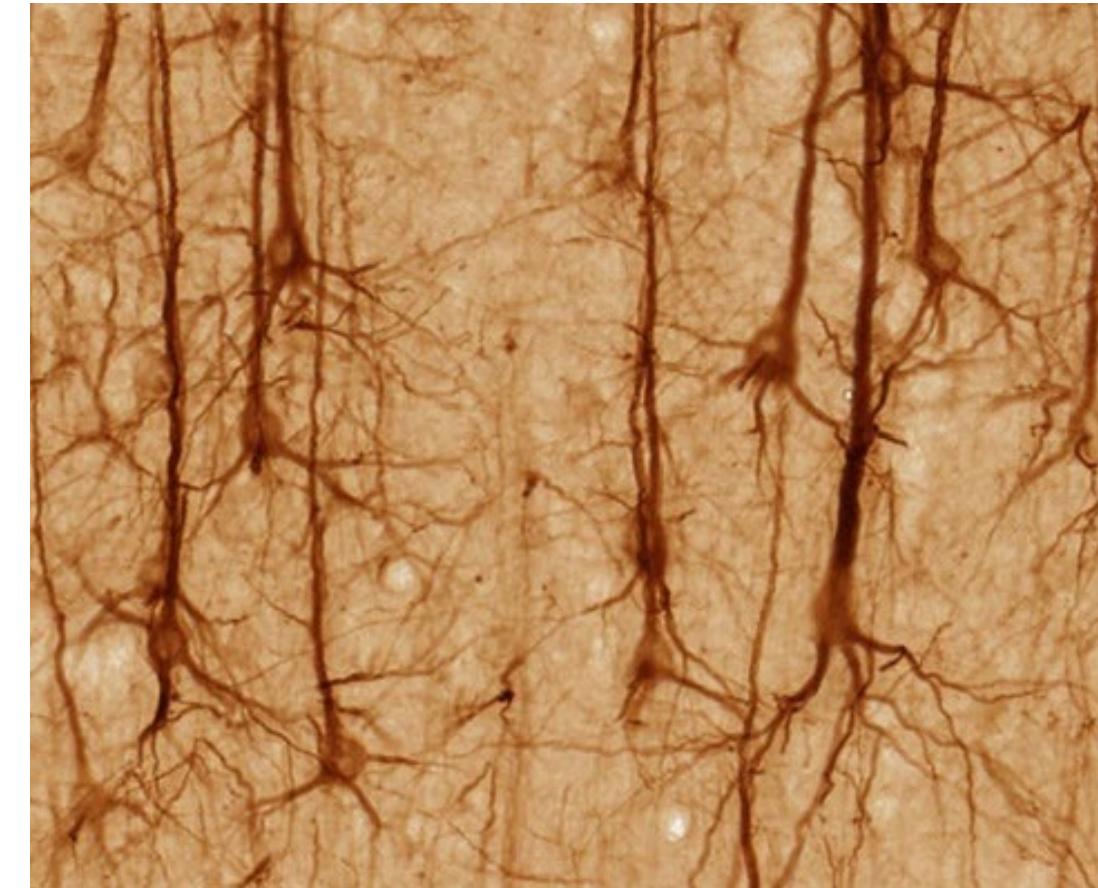
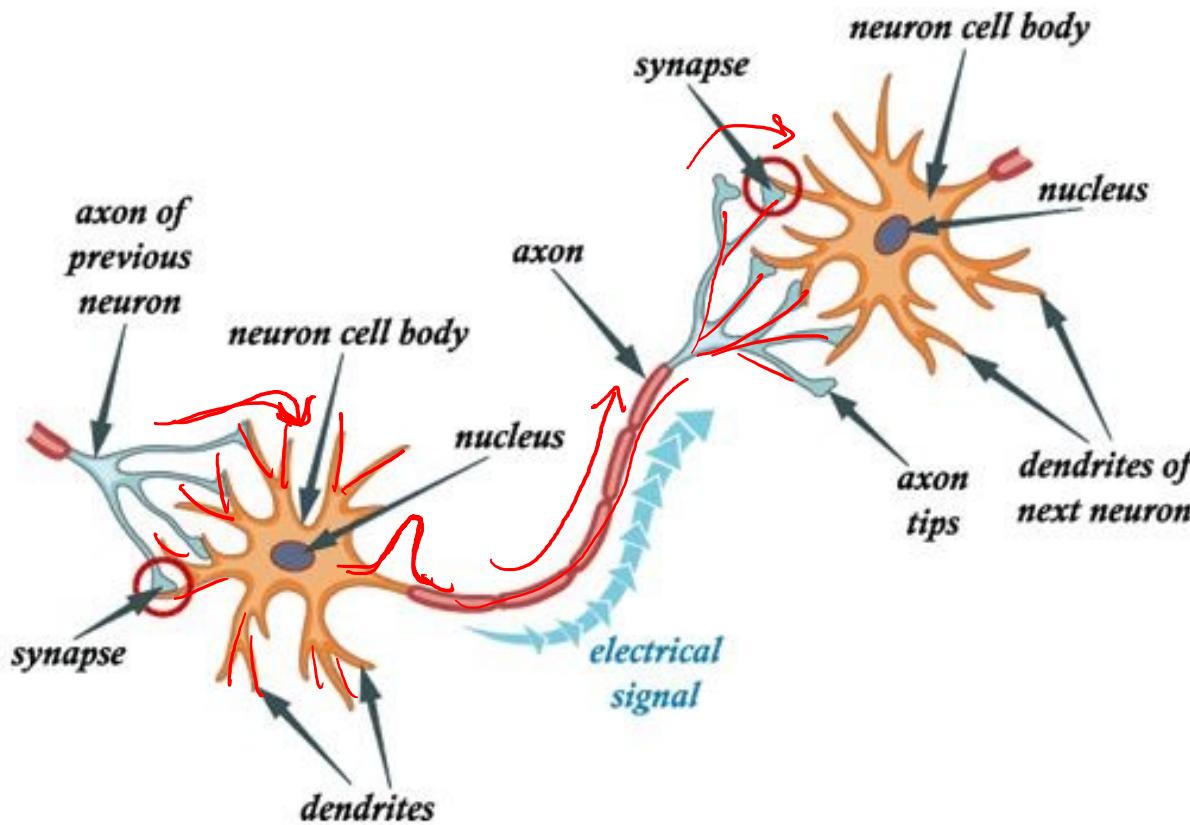
"young girl in pink shirt is swinging on swing."



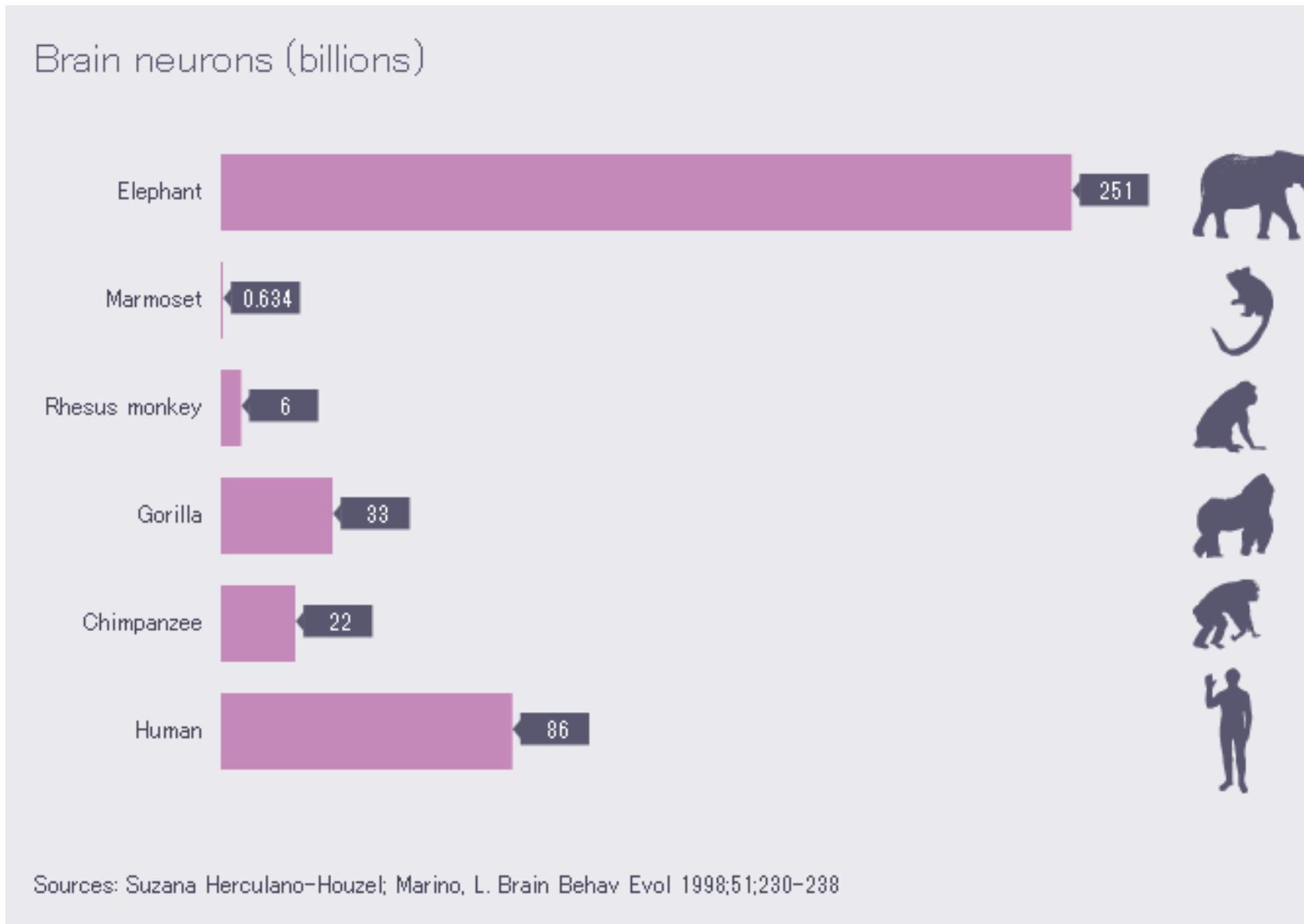
"man in blue wetsuit is surfing on wave."

The Deep Learning Age

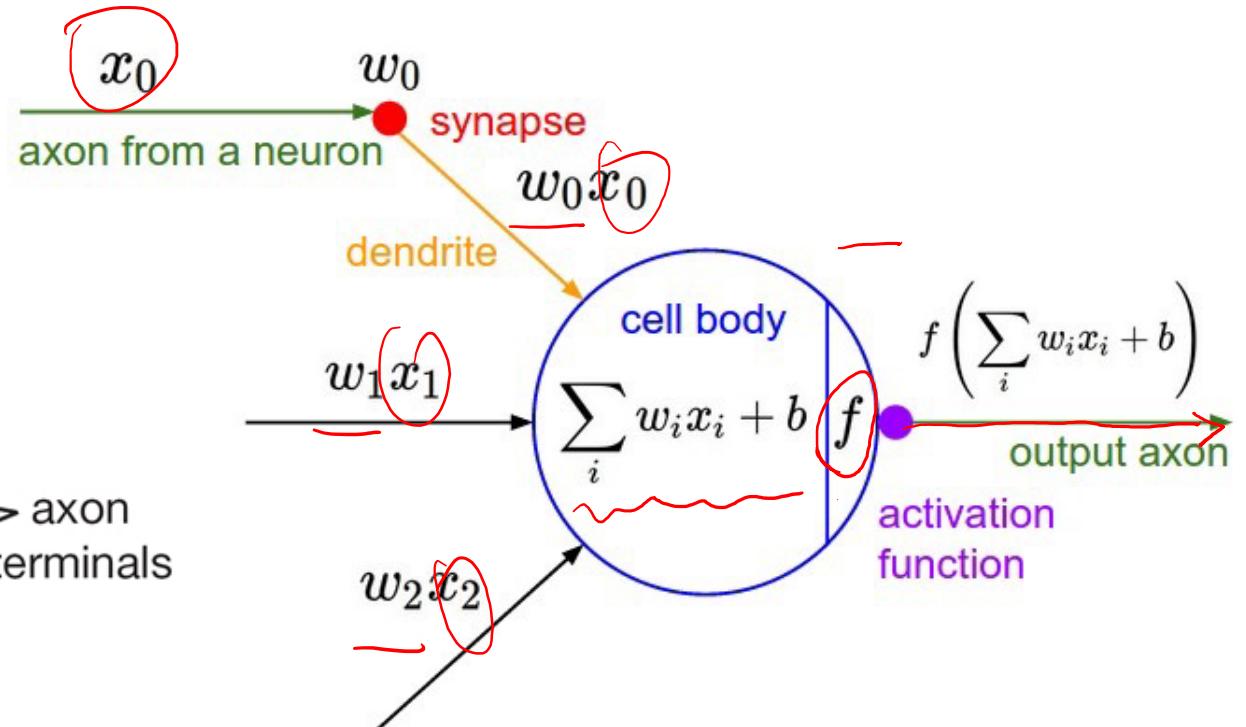
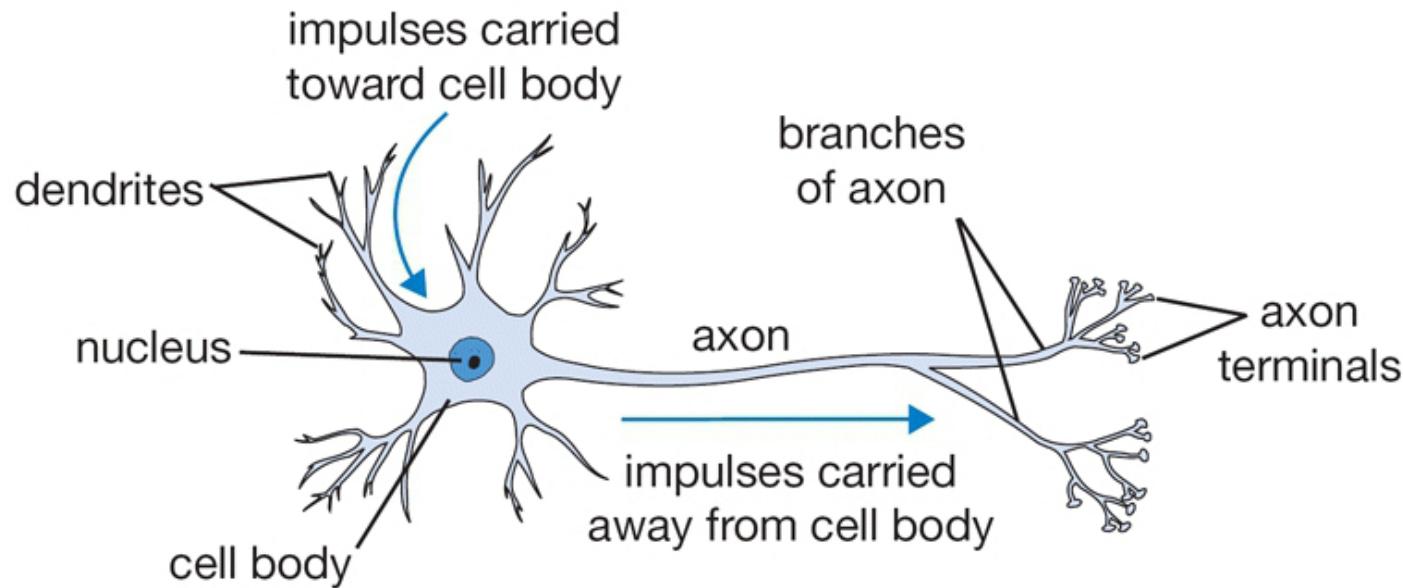
Neurons in Biological Systems



Neurons in Biological Systems

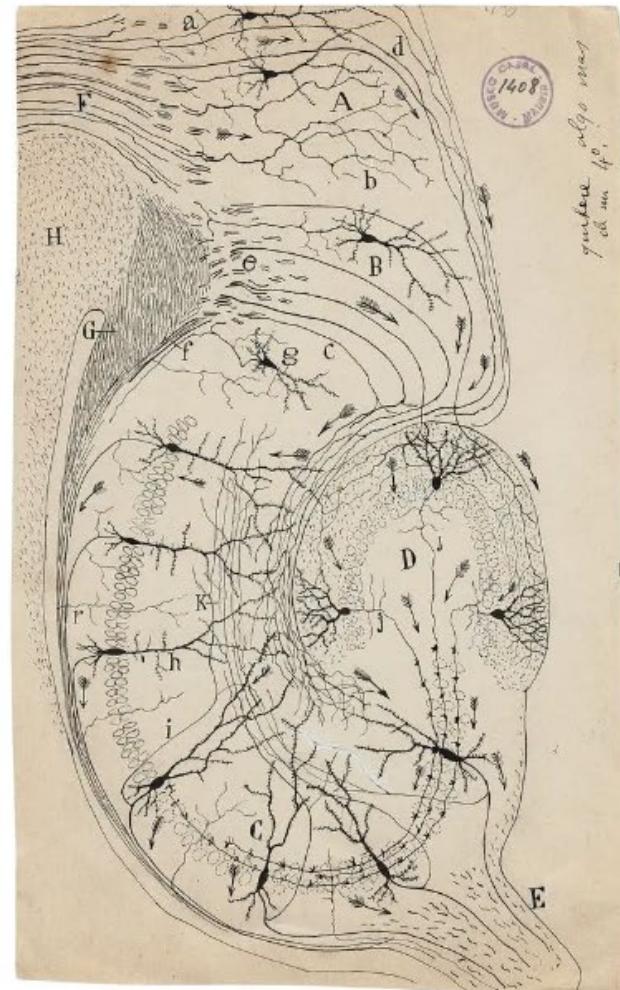
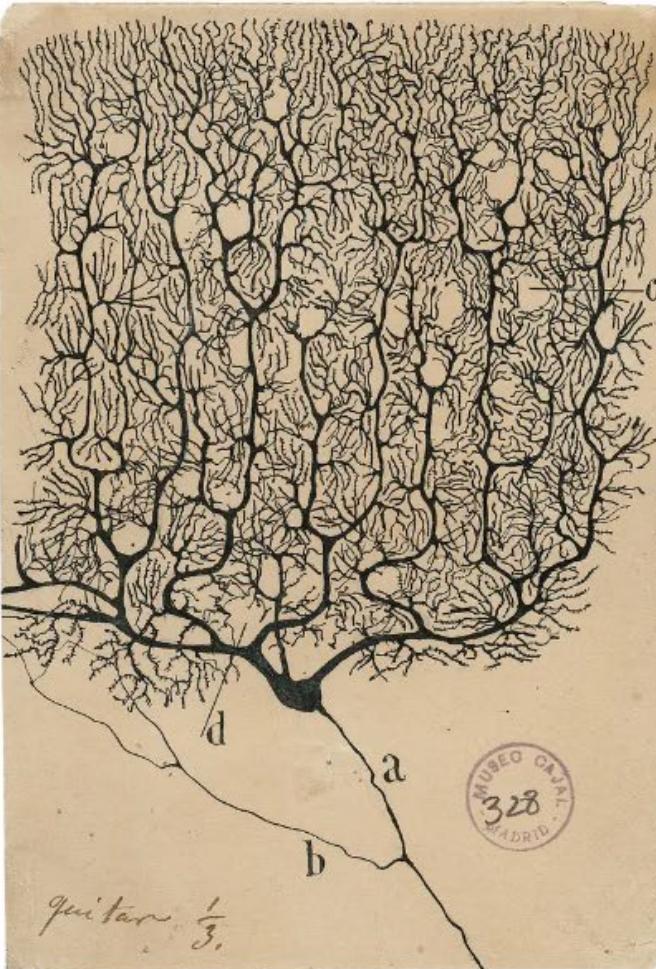
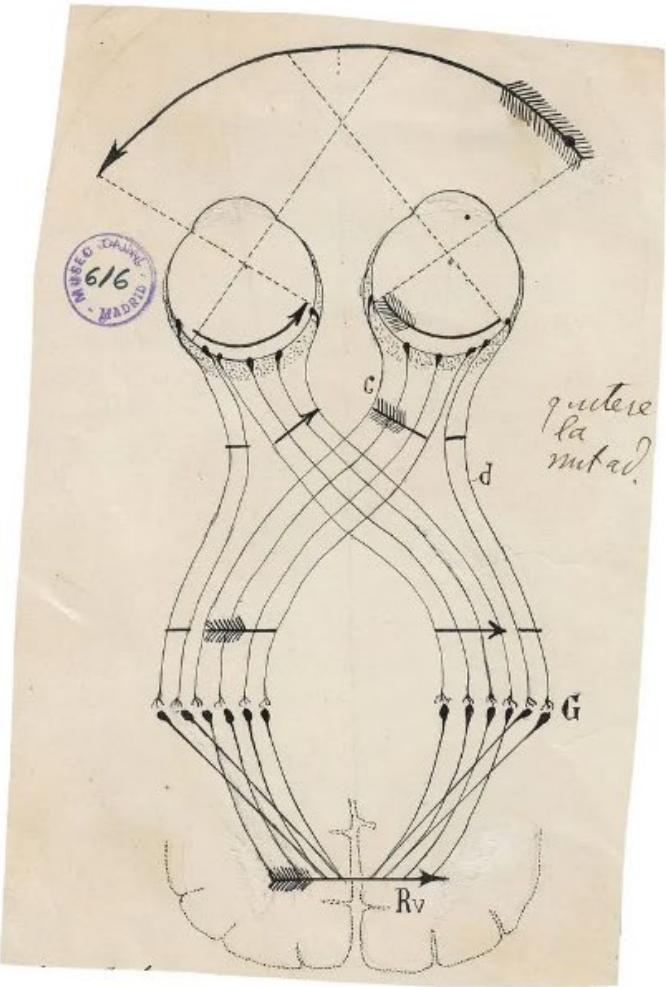


What is Artificial Neuron (Perceptron)



Perceptron

History of neural networks



Illustrations by Santiago Ramón y Cajal, the Spanish neuroscientist, from the book "The Beautiful Brain." From left: A diagram suggesting how the eyes might transmit a unified picture of the world to the brain; a purkinje neuron from the human cerebellum; and a diagram showing the flow of information through the hippocampus in the brain.

History of neural networks

“Connectionism”

Connectionism is an approach in the field of cognitive science that hopes to explain mental phenomena using artificial neural networks.
-Wikipedia



Alexander Bain: He worked on modeling of brain and how memory works using neural networks. *Mind and body* (1873)

History of neural networks

“Perceptron”



- Frank Rosenblatt
 - Psychologist, Logician
 - Inventor of the solution to everything, aka the Perceptron (1958)

History of neural networks

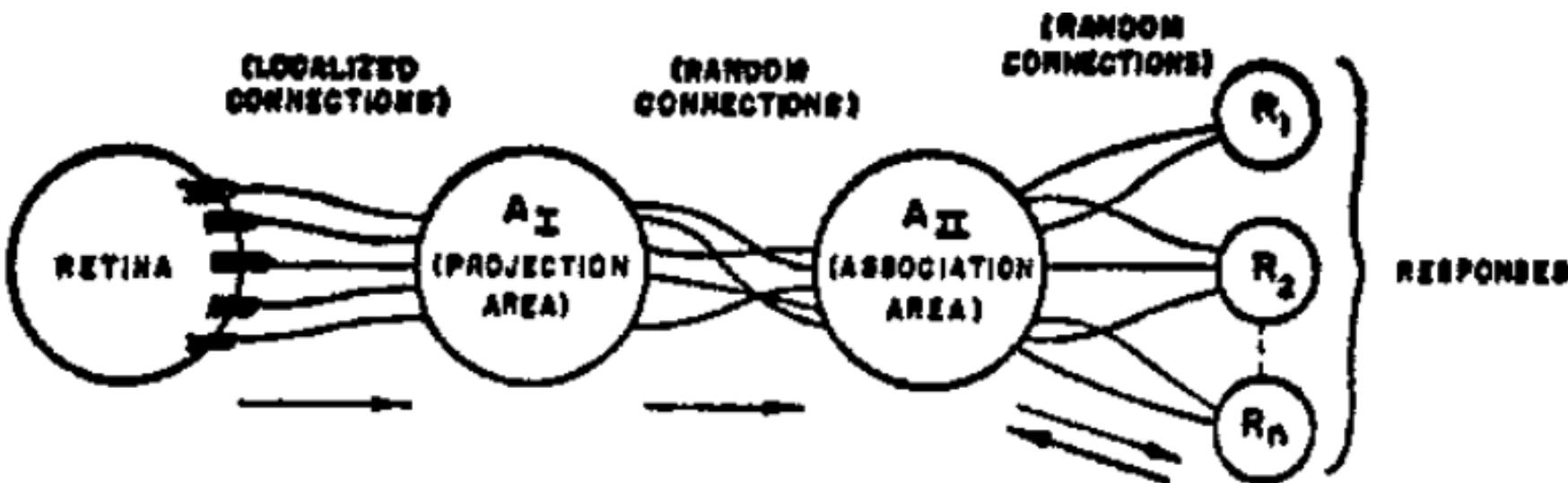
“Perceptron”

Psychological Review
Vol. 65, No. 6, 1958

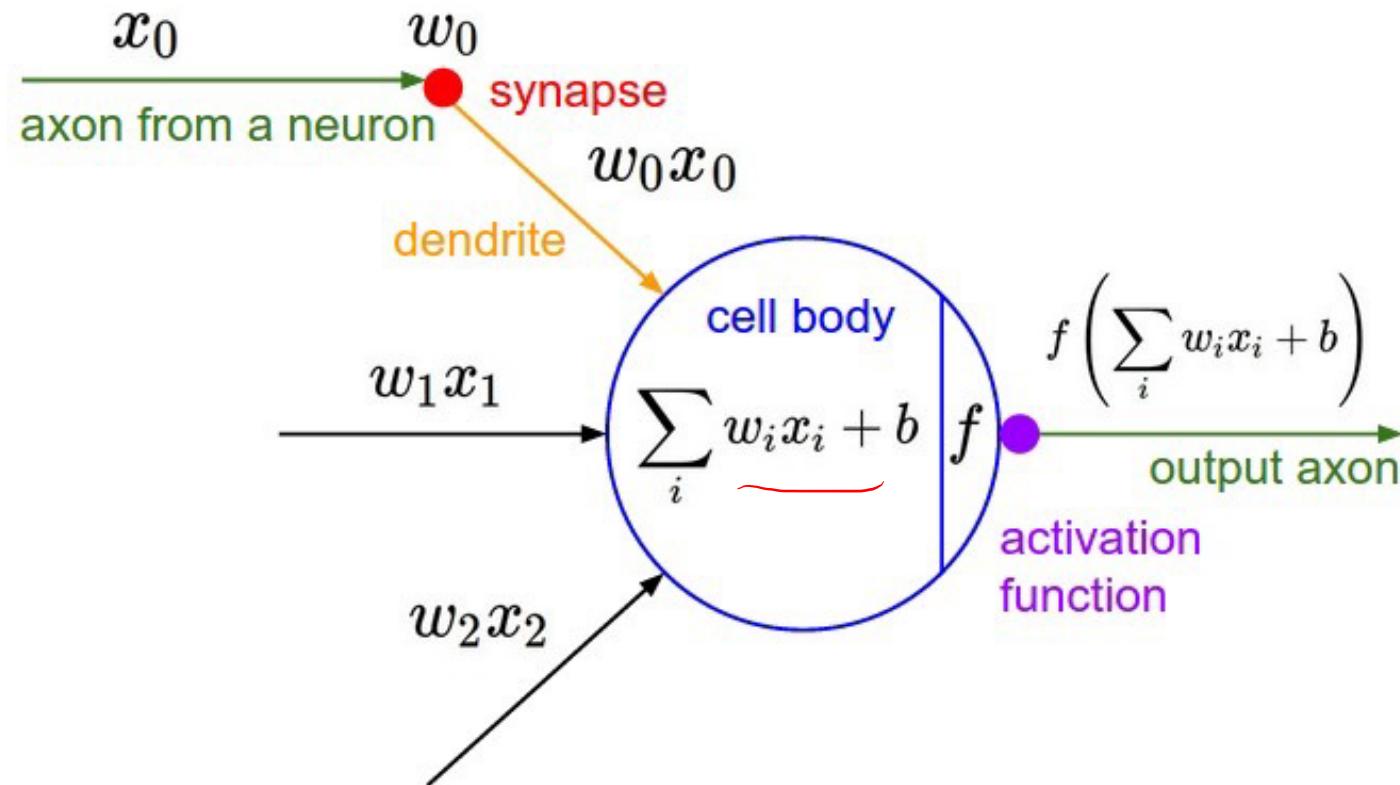
THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN¹

F. ROSENBLATT

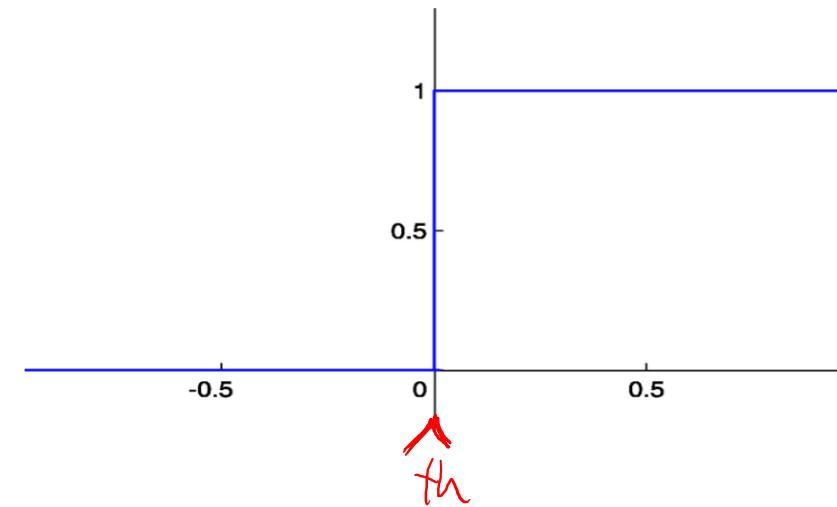
Cornell Aeronautical Laboratory



Perceptron

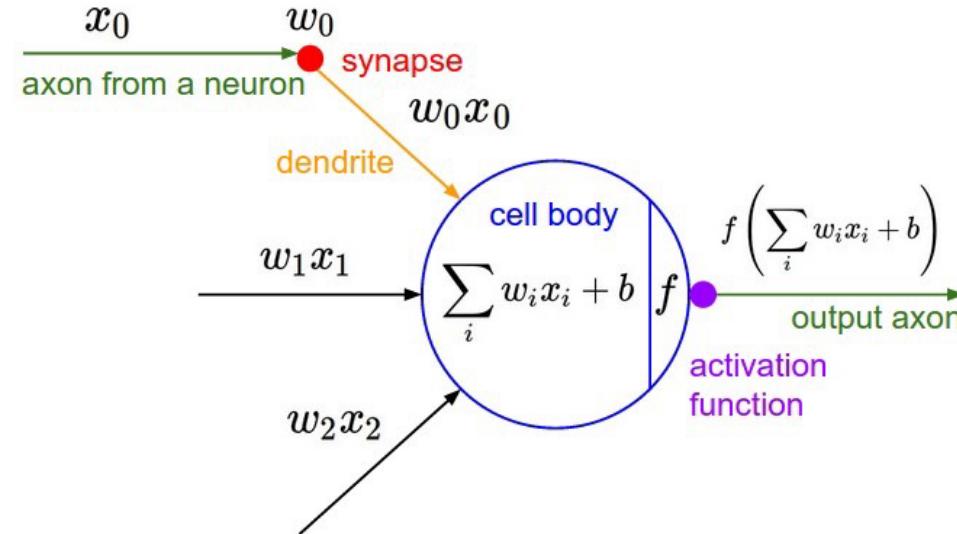


- Binary Threshold (Step function)

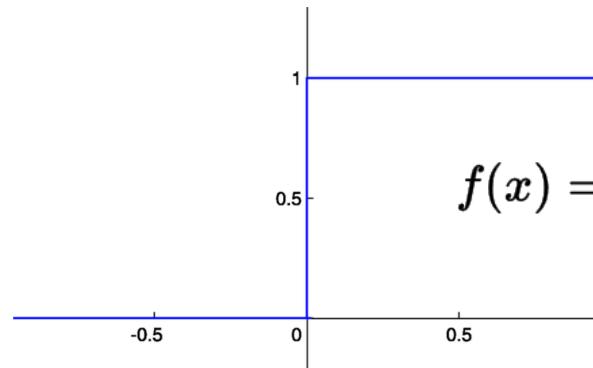


$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

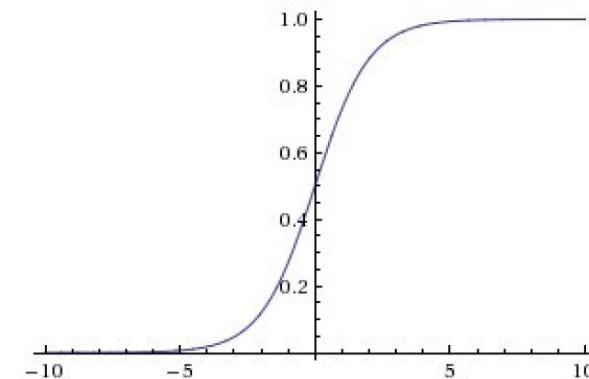
Activation functions



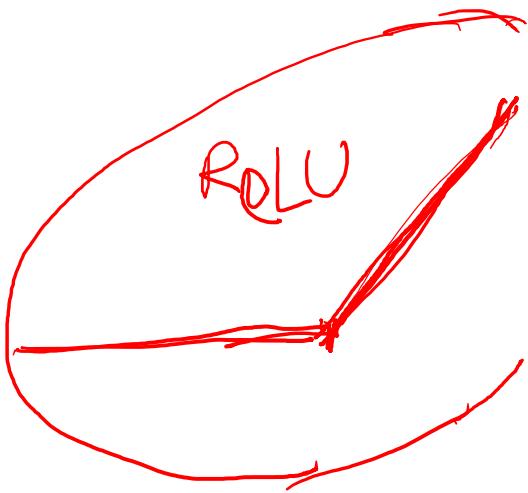
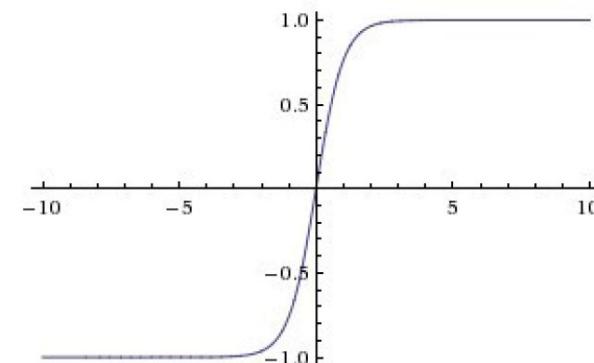
- Binary Threshold (Step function)



- Sigmoid $0 - 1$



- Tanh $-1 \sim 1$



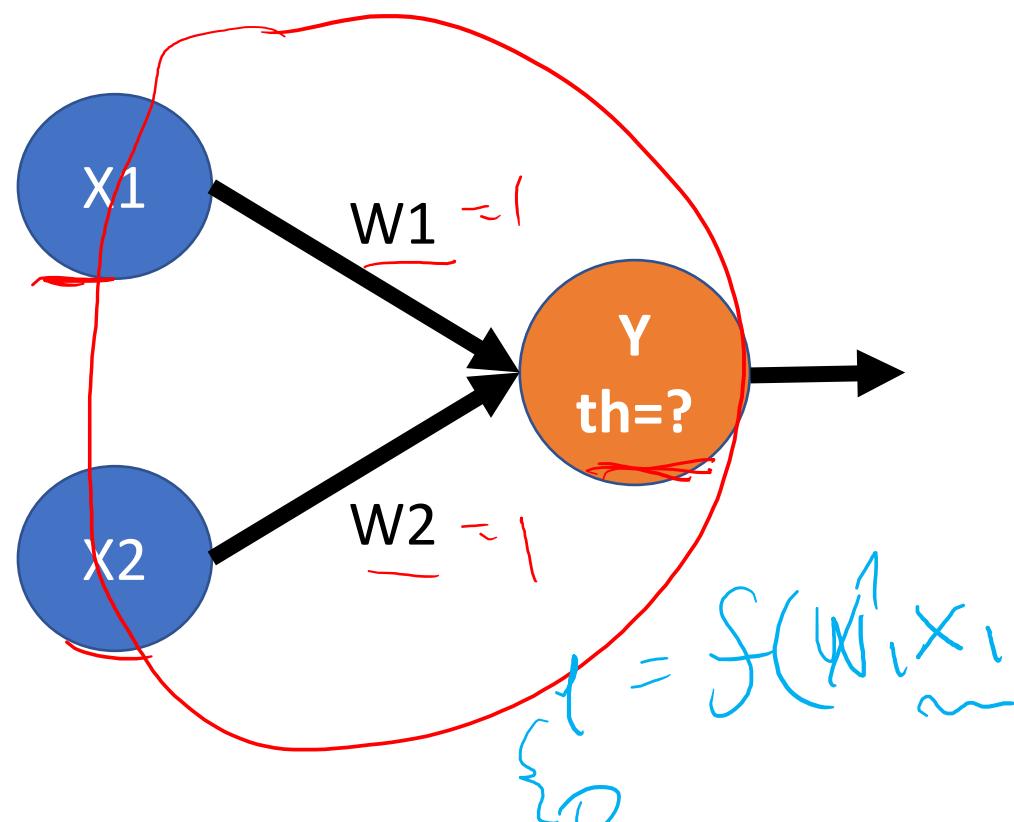
Perceptron

What can a perceptron do?

- Boolean tasks
- Update the weights whenever the perceptron output is wrong
- Proved convergence for linearly separable classes

Perceptron- quiz

A perceptron does not need to have binary input, but let's consider a perceptron as a Boolean logic gate



X_1 and X_2 are binary inputs (either 0 or 1)
 W_1 and W_2 are weights set to 1
What should the step activation function threshold “th” in order to have AND gate output?

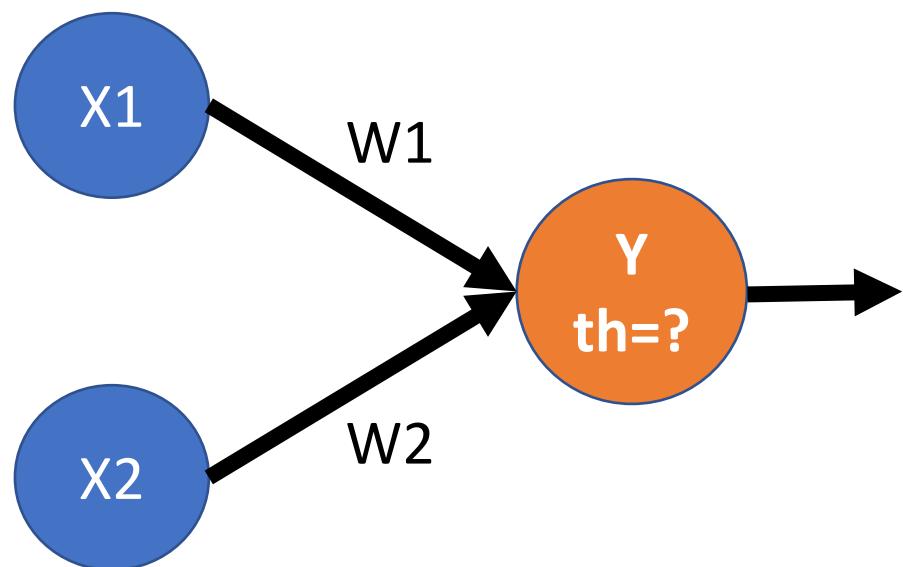
$$y = S(w_1x_1 + w_2x_2)$$
$$\begin{array}{cccc} x_1 & x_2 & \text{th} & y \\ 1 & 1 & 2 & 1 \end{array}$$

0	1
1	0

0	0
0	0

Perceptron- quiz

A perceptron does not need to have binary input, but let's consider a perceptron as a Boolean logic gate



X_1 and X_2 are binary inputs (either 0 or 1)
 w_1 and w_2 are weights set to 1
What should the step activation function threshold “th” in order to have OR gate output?

A handwritten truth table for an OR gate. The columns are labeled X_1 , X_2 , th , and Y . The rows show the following values:

X_1	X_2	th	Y
1	0	1	1
0	1	1	1
1	1	2	1

Below the table, there is handwritten text: $1 + 0 = 1$, $0 + 1 = 1$, and $1 + 1 = 2$.

Perceptron

A perceptron can do more than just a binary logic gate.
It can also do classification or regression tasks.

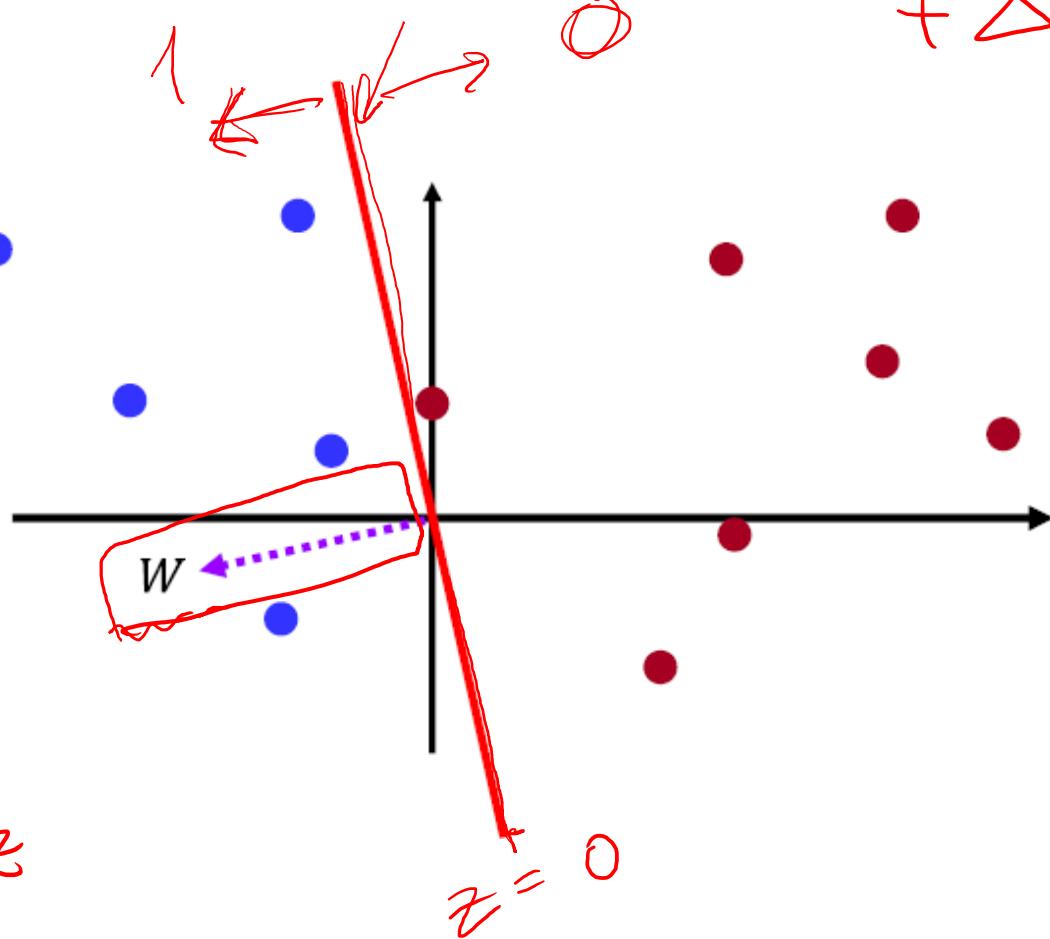
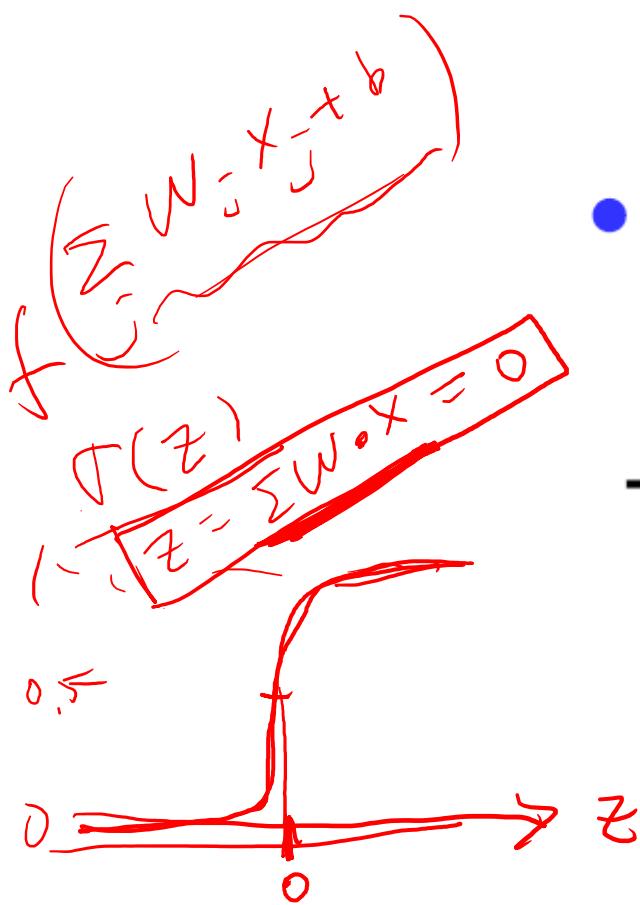
(1) (1)

Learning in perceptron

- Perceptron rule
- Delta rule (Gradient Descent)

Perceptron

Perceptron rule



$$w_j \leftarrow w_j^{\text{old}} - \alpha (\hat{y}_i - y_i) X_{ij}$$

Key: Red -1, Blue = +1

pred target

$+ \Delta w_j$

1
2
3
...
 n

Perceptron

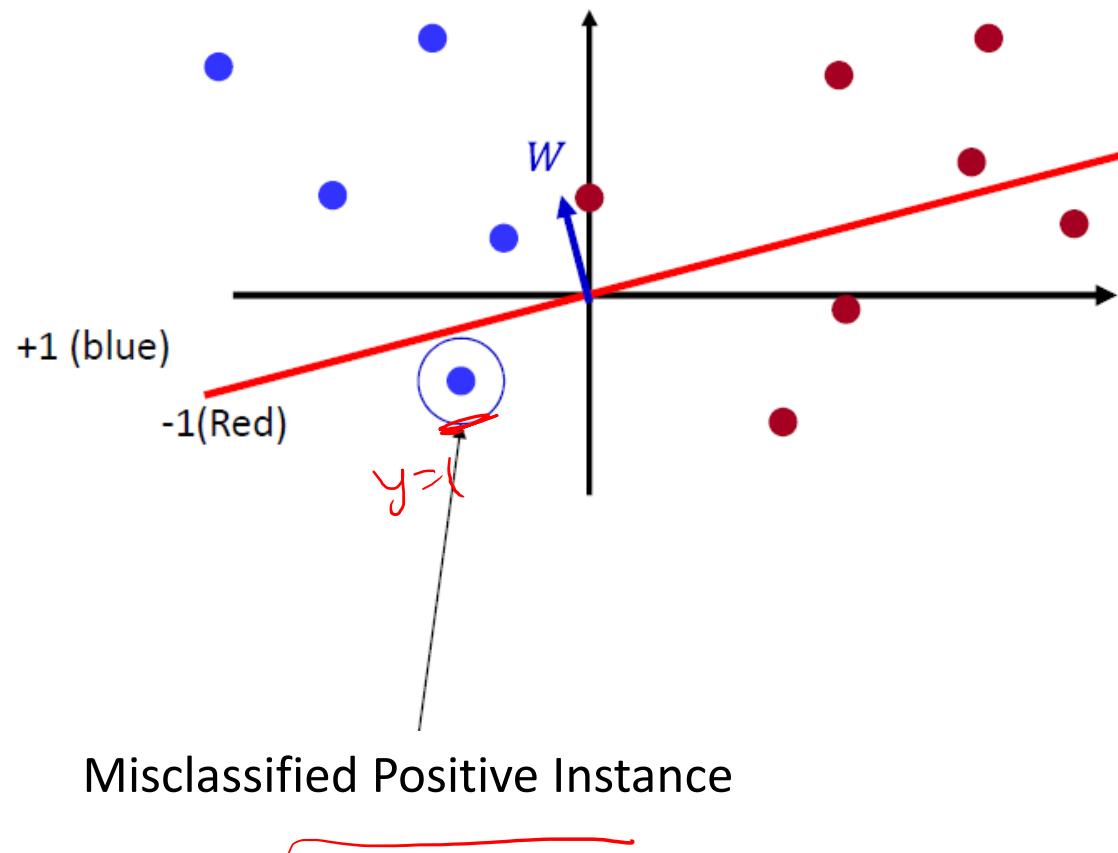
Perceptron algorithm

- Cycle through the training instances
- Only update W on misclassified instances
- If instance misclassified:
 - If instance is positive class (positive misclassified as negative)
 $y=1$
 $W = W + \underbrace{X_i}_{}$
 - If instance is negative class (negative misclassified as positive)
 $y=0$
 $W = W - \underbrace{X_i}_{}$

$$W = W - \underbrace{X_i}_{}$$

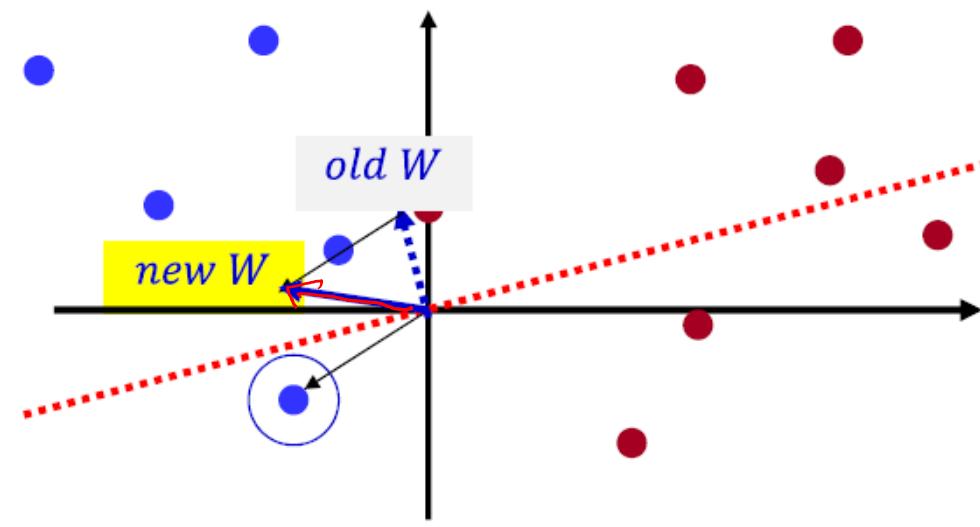
Perceptron

Perceptron algorithm



Perceptron

Perceptron algorithm

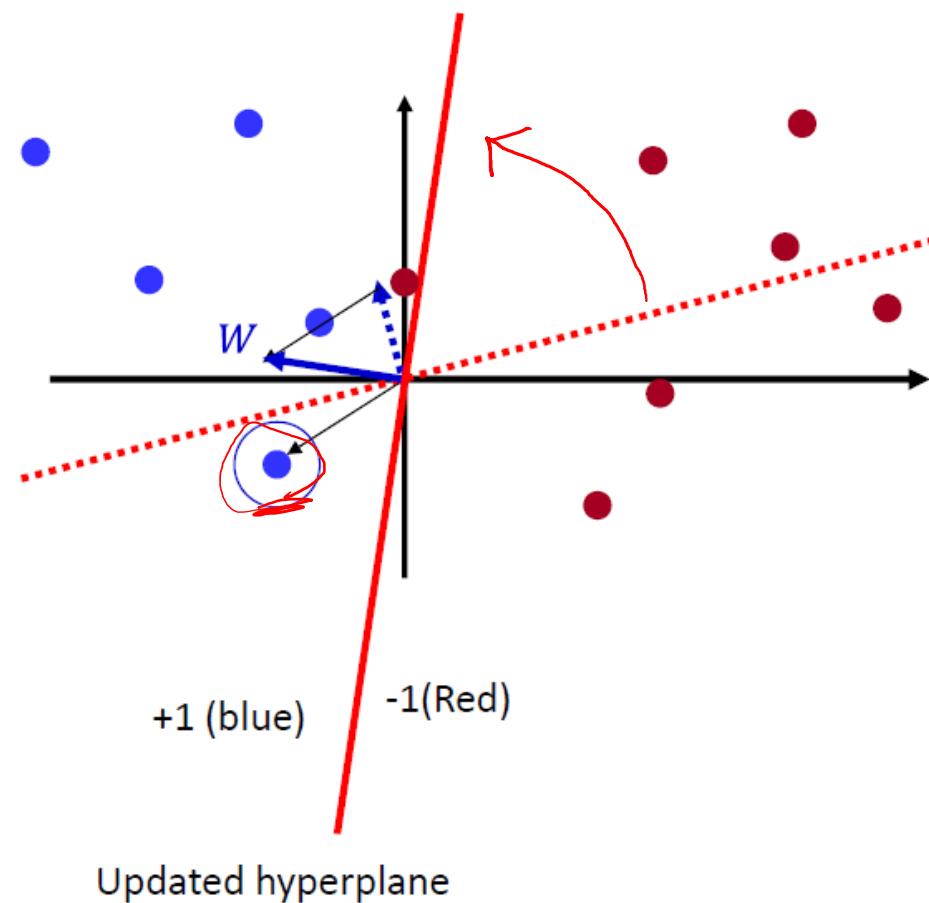


Updated weight vector

Update weight (add to W) for the misclassified positive instance

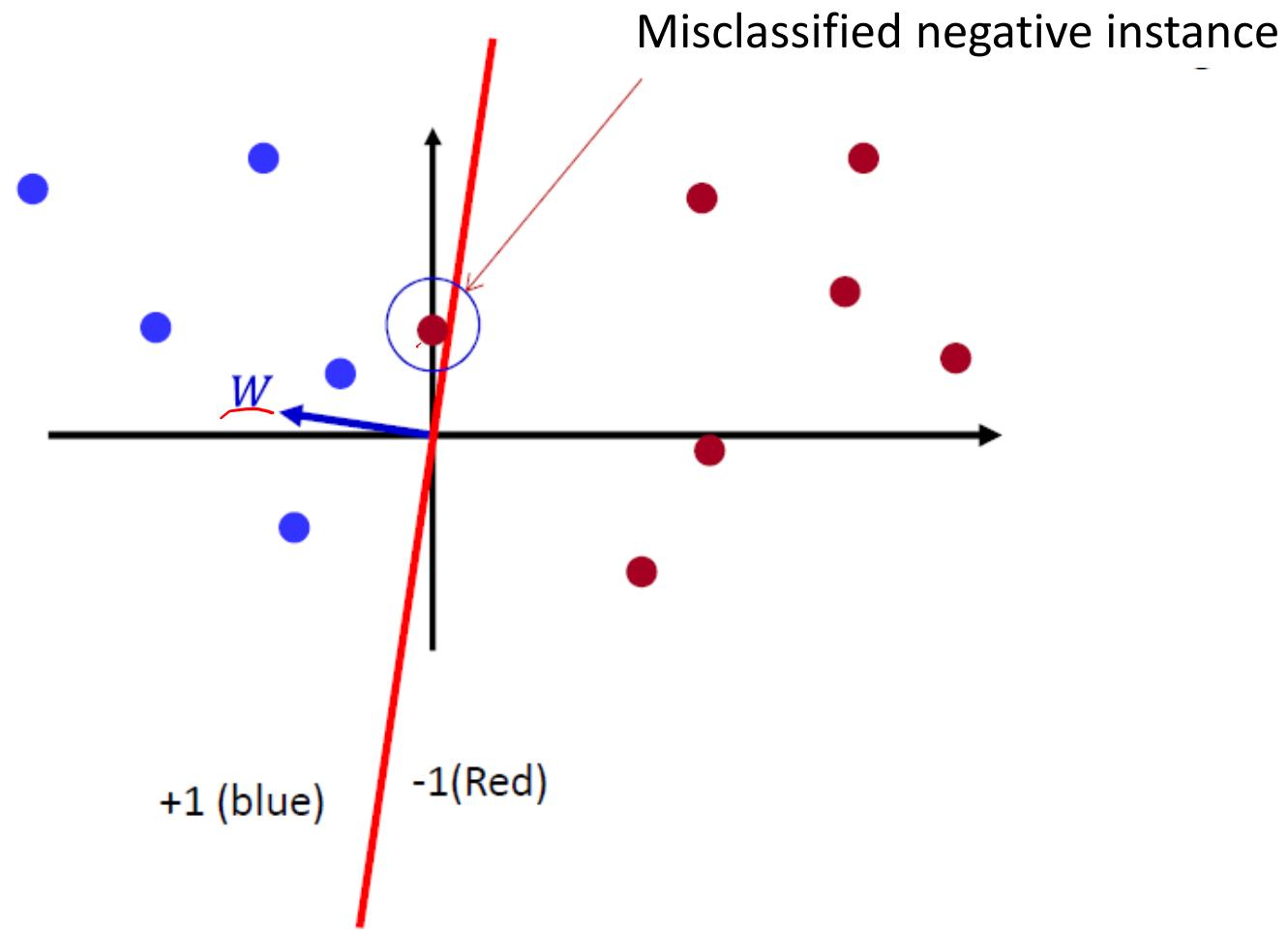
Perceptron

Perceptron algorithm



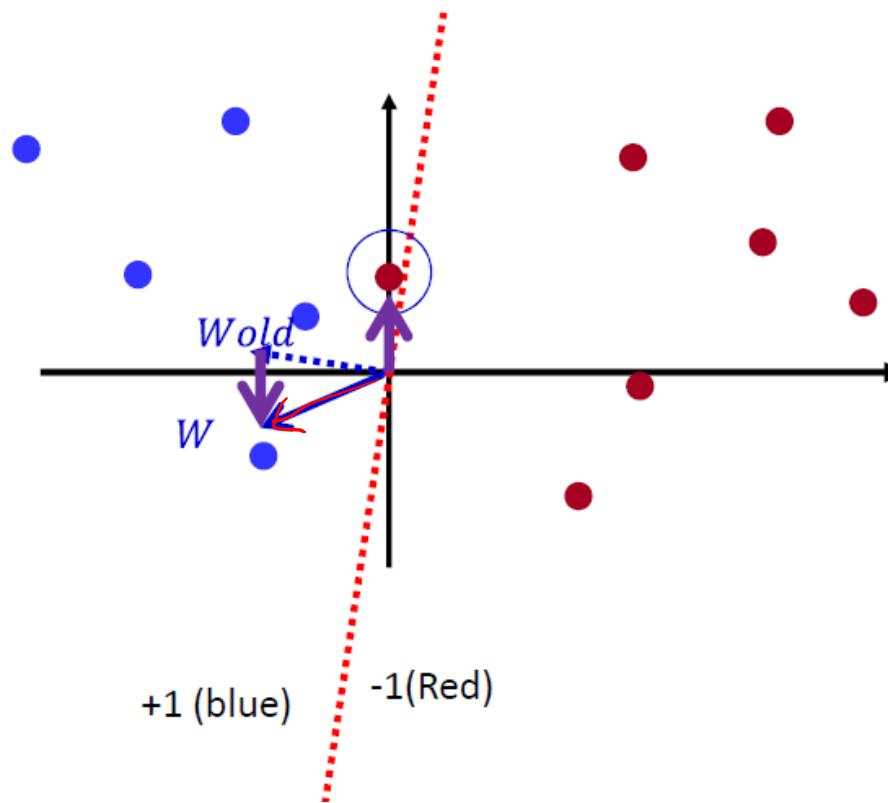
Perceptron

Perceptron algorithm



Perceptron

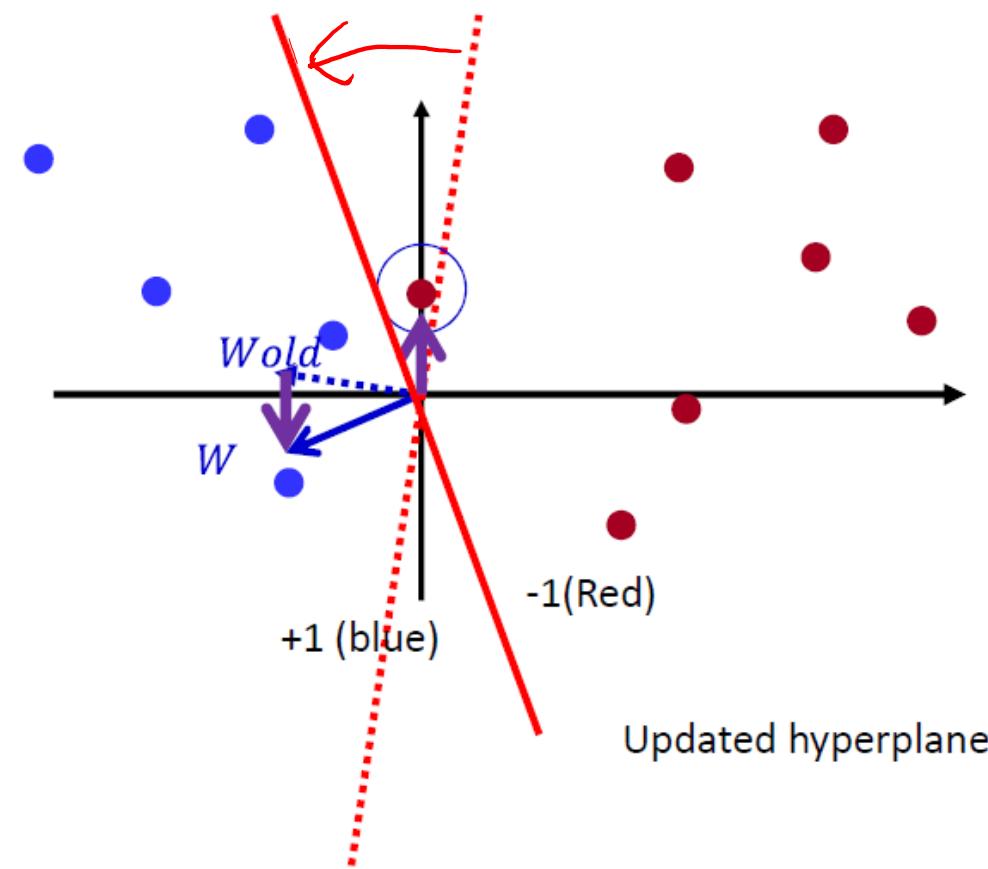
Perceptron algorithm



Update weight (subtract from W) for the misclassified negative instance

Perceptron

Perceptron algorithm



Perceptron

Delta rule (Gradient Descent)

$$\omega_j \leftarrow \omega_j + \frac{\Delta \omega_j}{\text{MSE}}$$

$$\omega_j \leftarrow \omega_j - \alpha \frac{\partial \mathcal{L}}{\partial \omega_j}$$

$$\mathcal{L} = \underbrace{\frac{1}{2}(\hat{y}_i - y_i)^2}_{\text{MSE}}$$

$$\text{predict } \hat{y}_i = f\left(\sum_j \omega_j X_{ij}\right)$$

$$\omega_j \leftarrow \omega_j - \alpha (\hat{y}_i - y_i) X_{ij}$$

Perceptron

Learning in perceptron- conclusion

- Perceptron rule

$$\omega_j \leftarrow \omega_j - \underbrace{\alpha}_{\text{learning rate}} (\hat{y}_i - y_i) X_{ij}$$

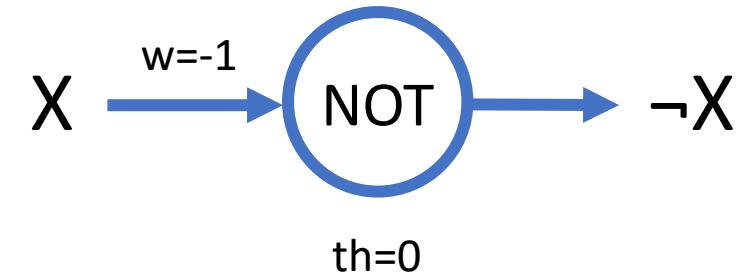
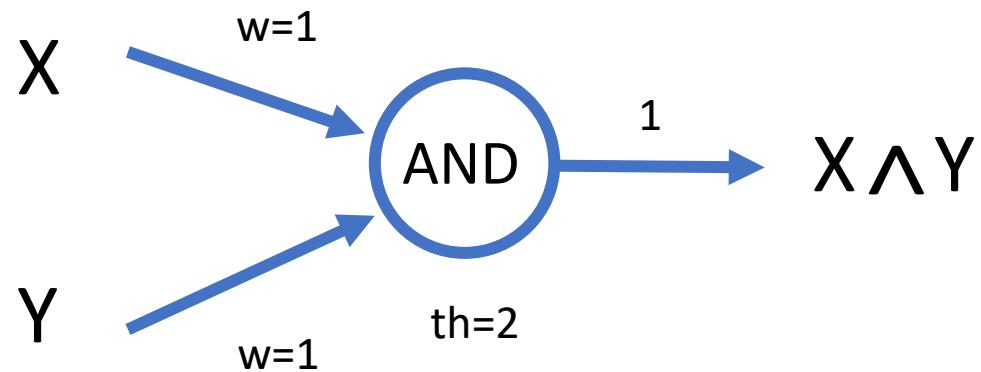
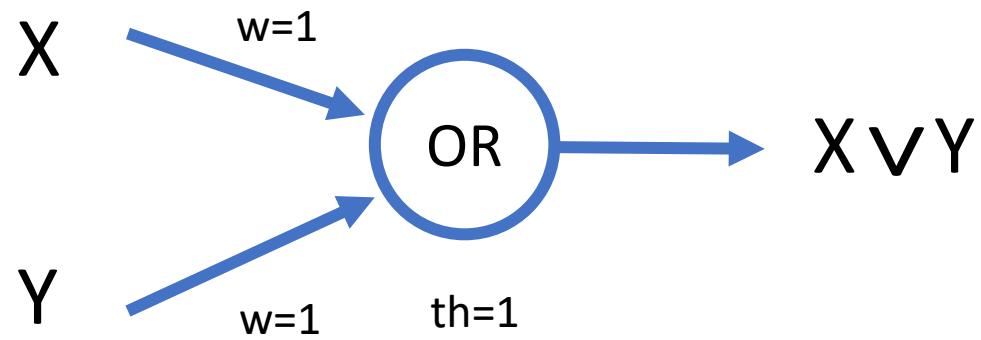
- Delta rule (Gradient Descent)

$$\omega_j \leftarrow \omega_j - \underbrace{\alpha}_{\text{stepsize}} (\hat{y}_i - y_i) X_{ij}$$

learning rate
stepsize

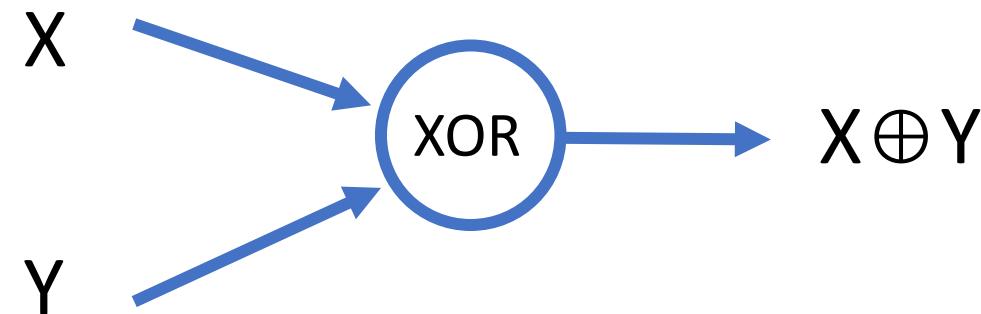
$\alpha < 1$

Perceptron



Perceptron

Not possible to determine weights and threshold for XOR gate!

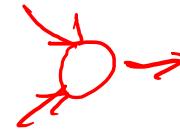
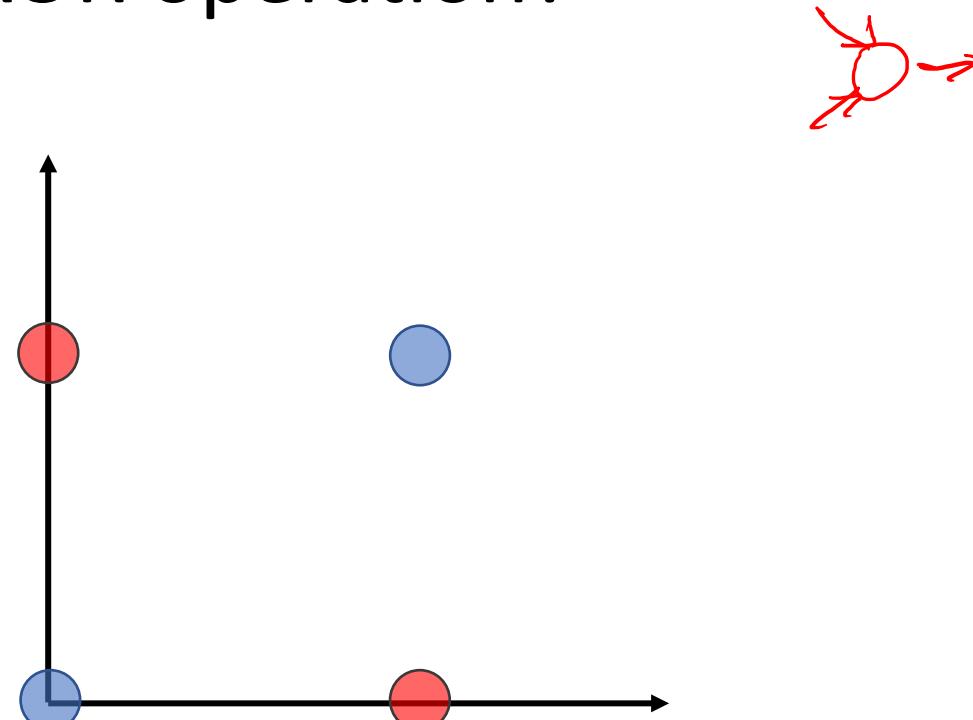


- Minsky and Papert, 1968

Perceptron

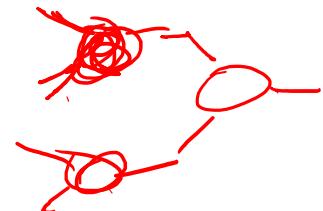
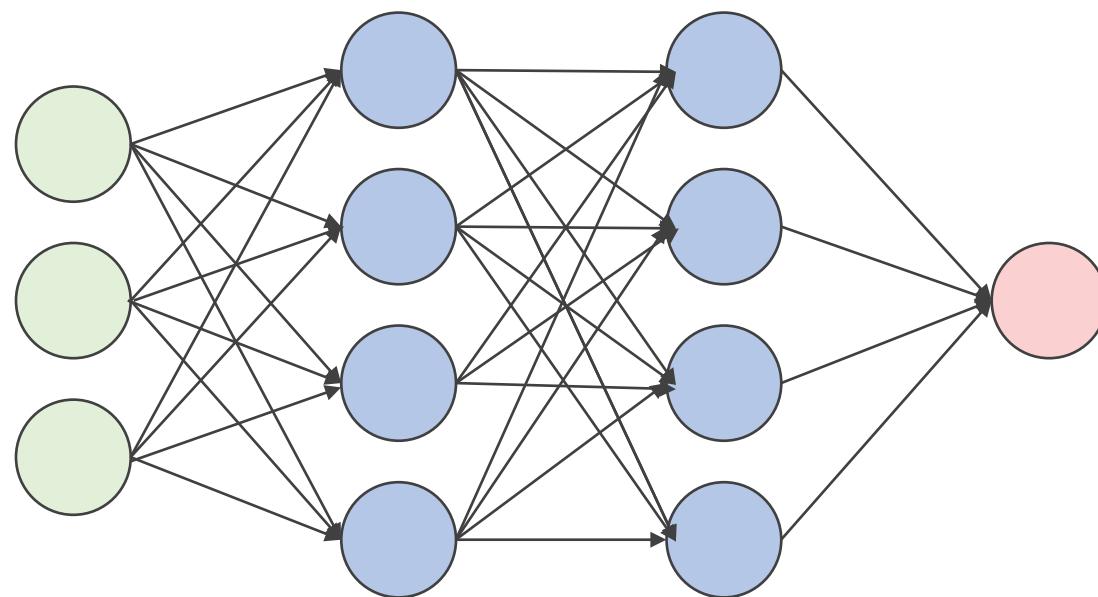
What's special about XOR operation?

INPUT		OUTPUT
A	B	A XOR B
1	1	0
0	0	0
0	1	1
1	0	1
1	1	0

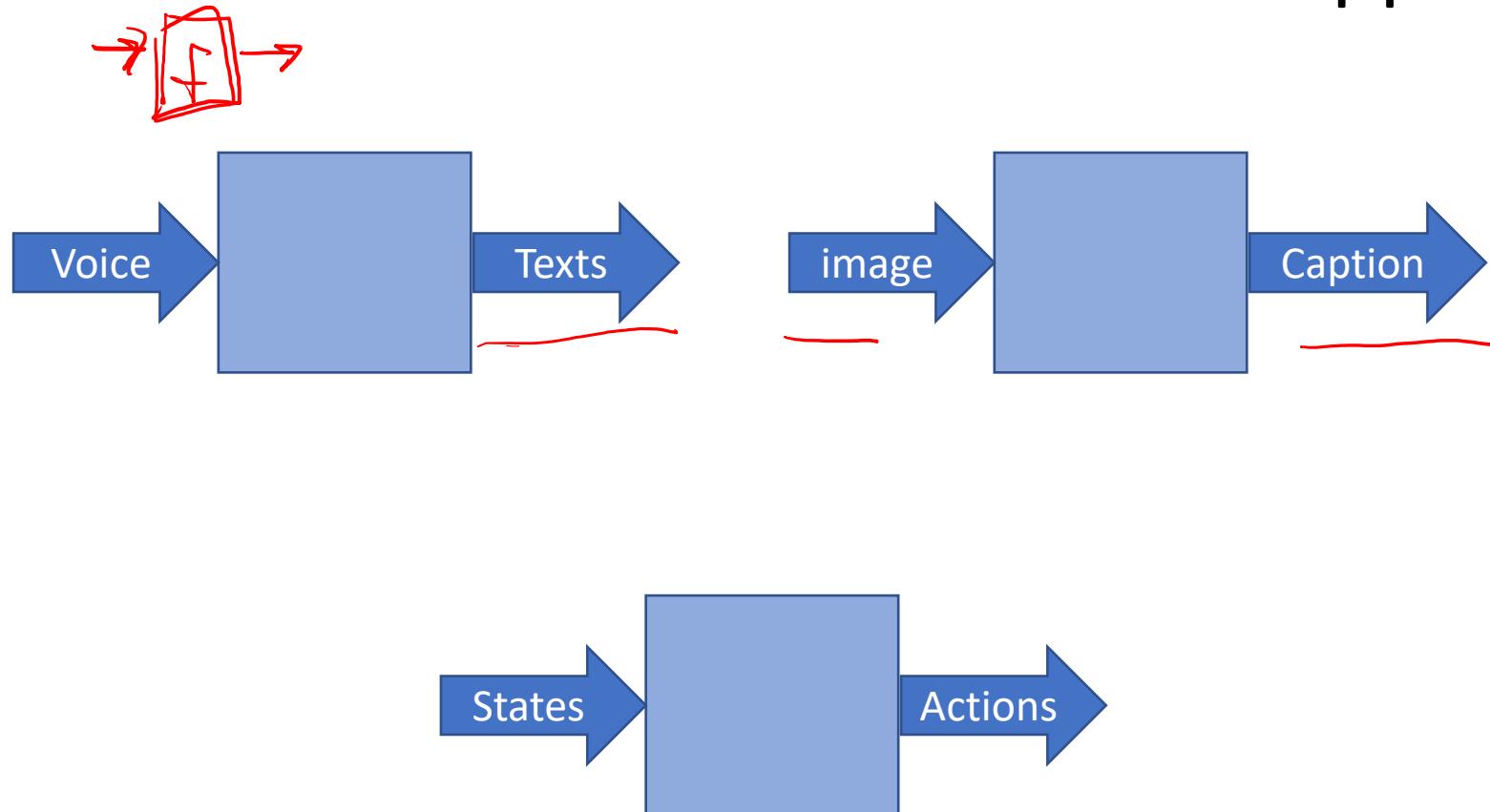


How do we handle linearly inseparable cases?

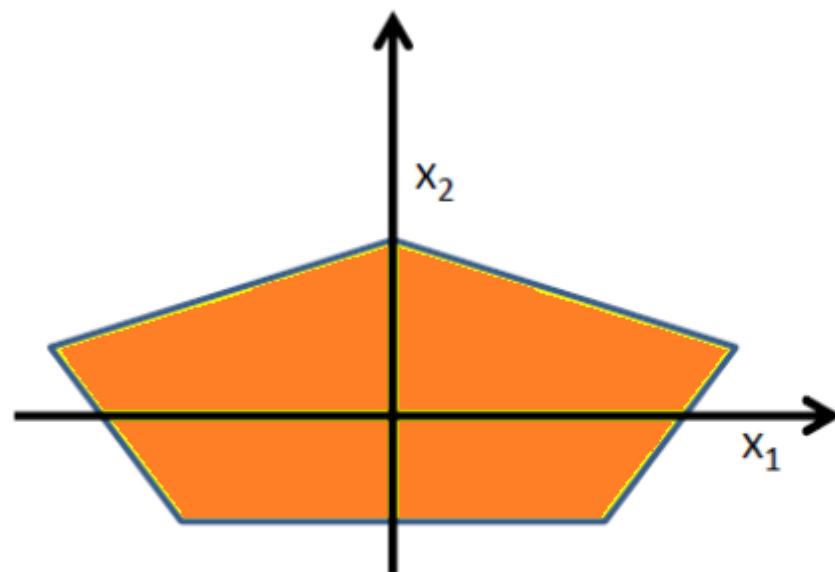
How do we handle linearly inseparable cases?



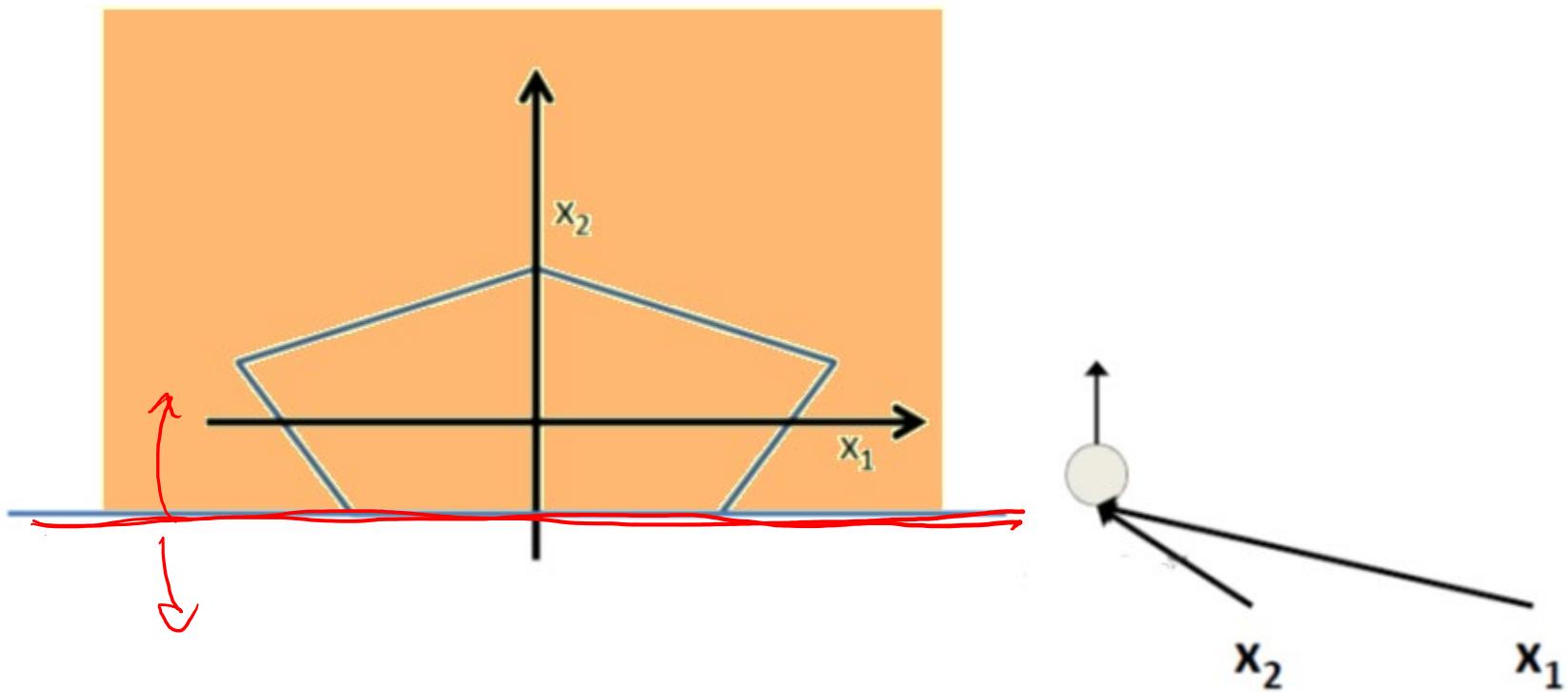
Neural networks are universal function approximator



We can make an arbitrary shape decision boundary

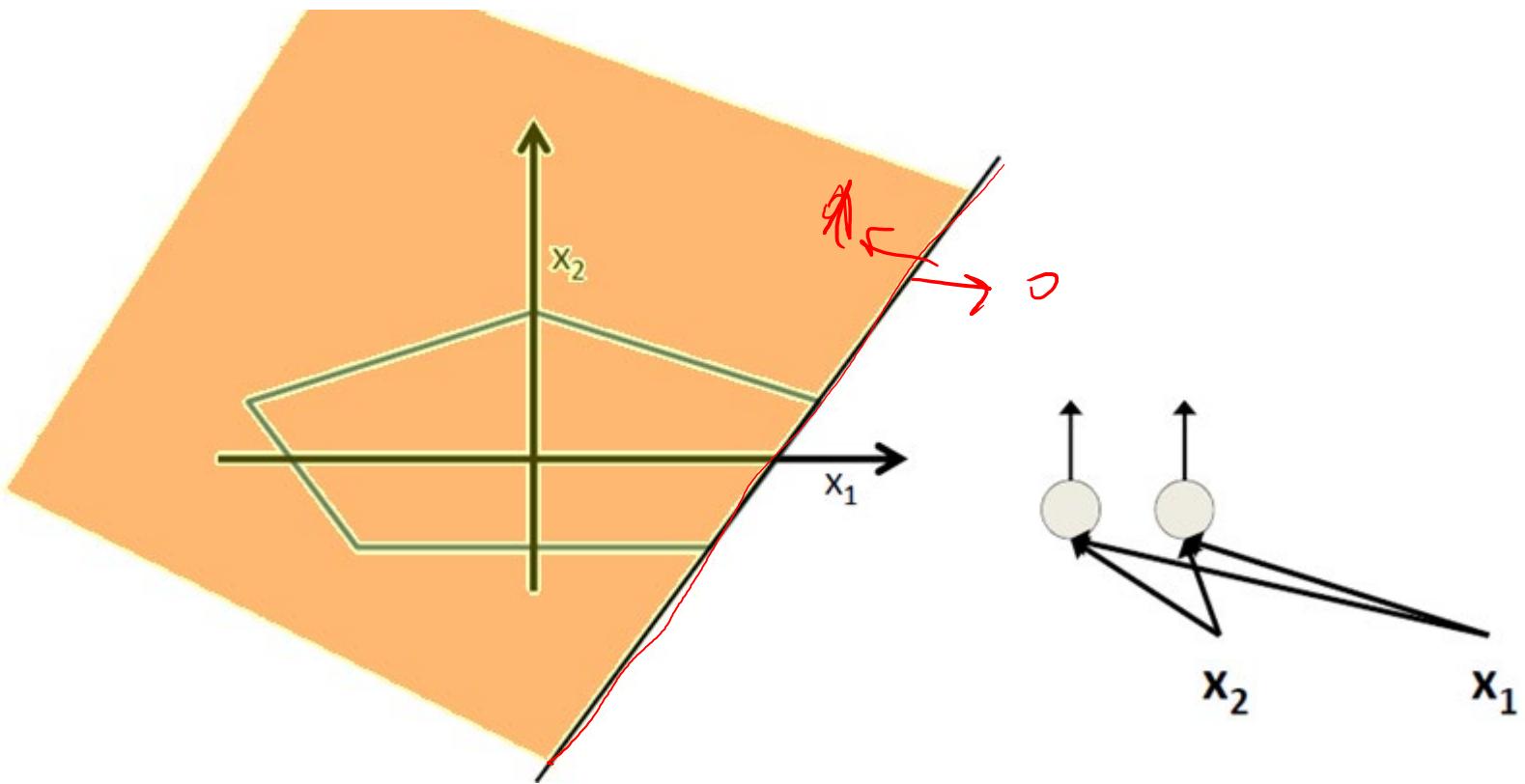


MLP

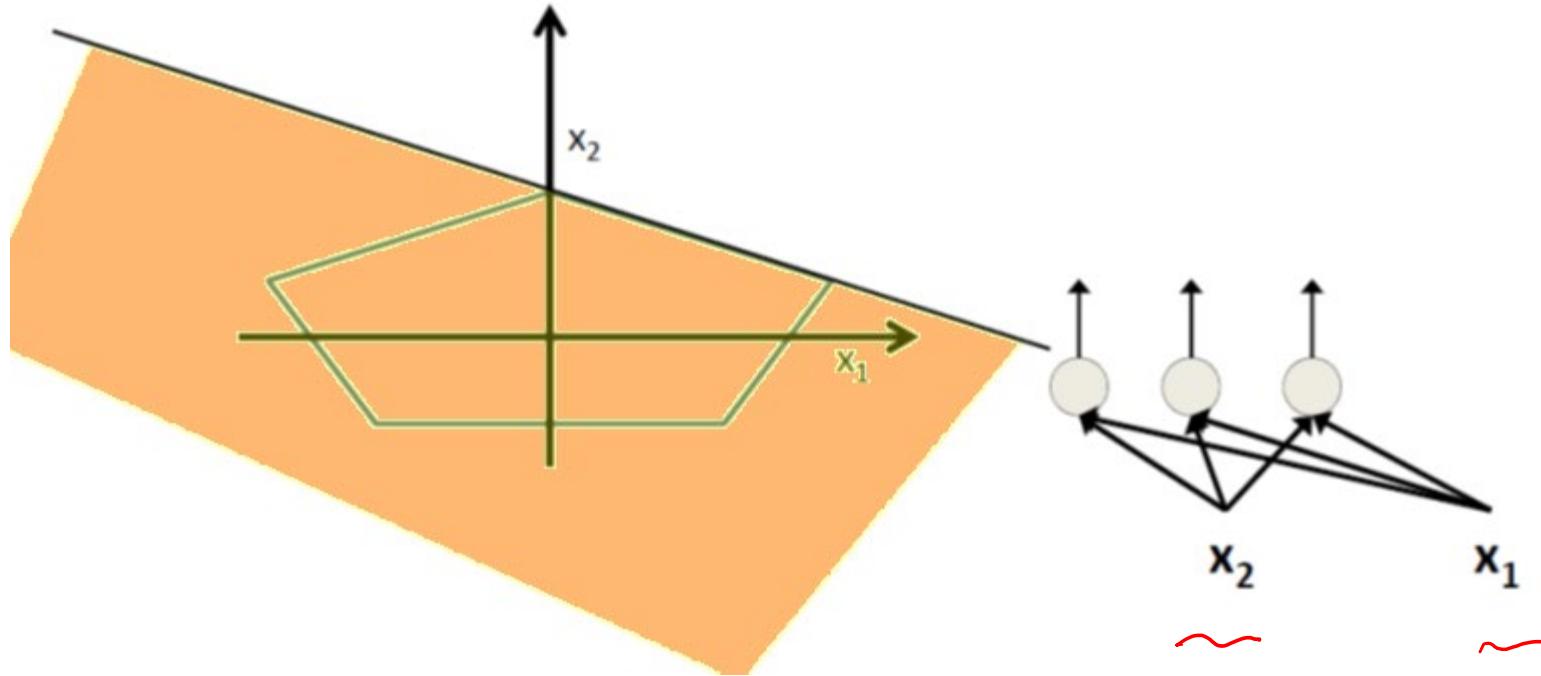


The colored area means that the neurons activate in the region

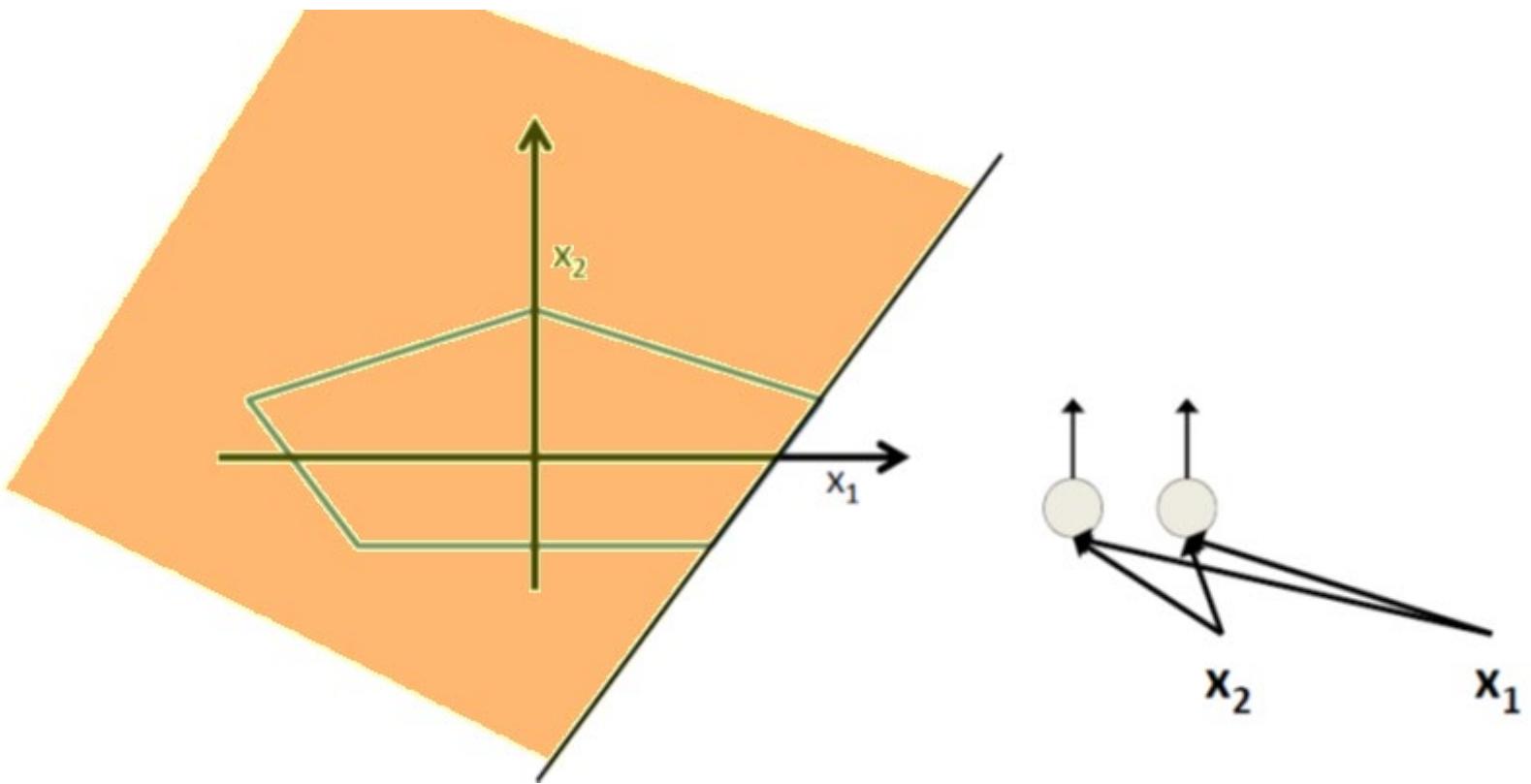
MLP



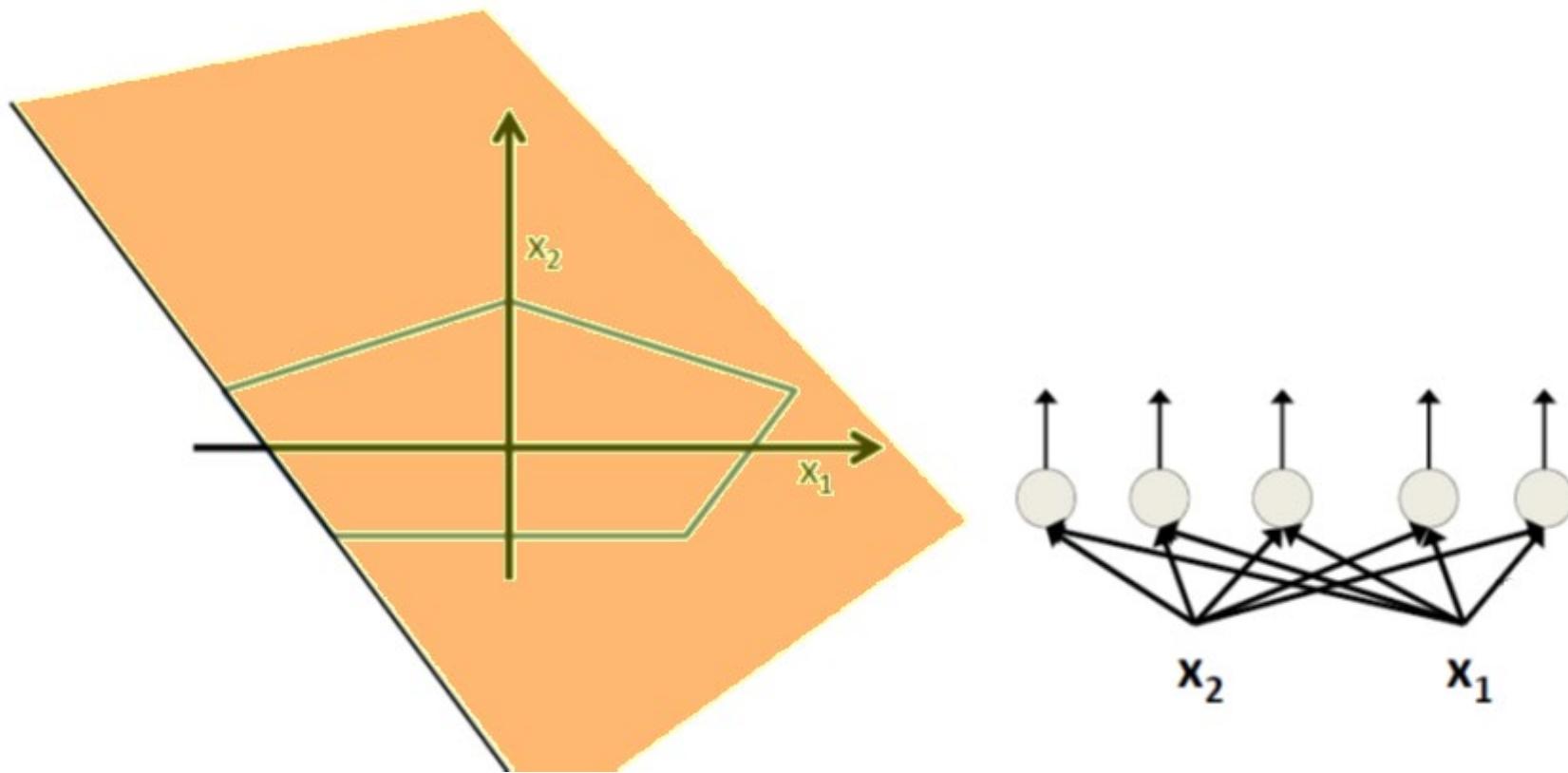
MLP



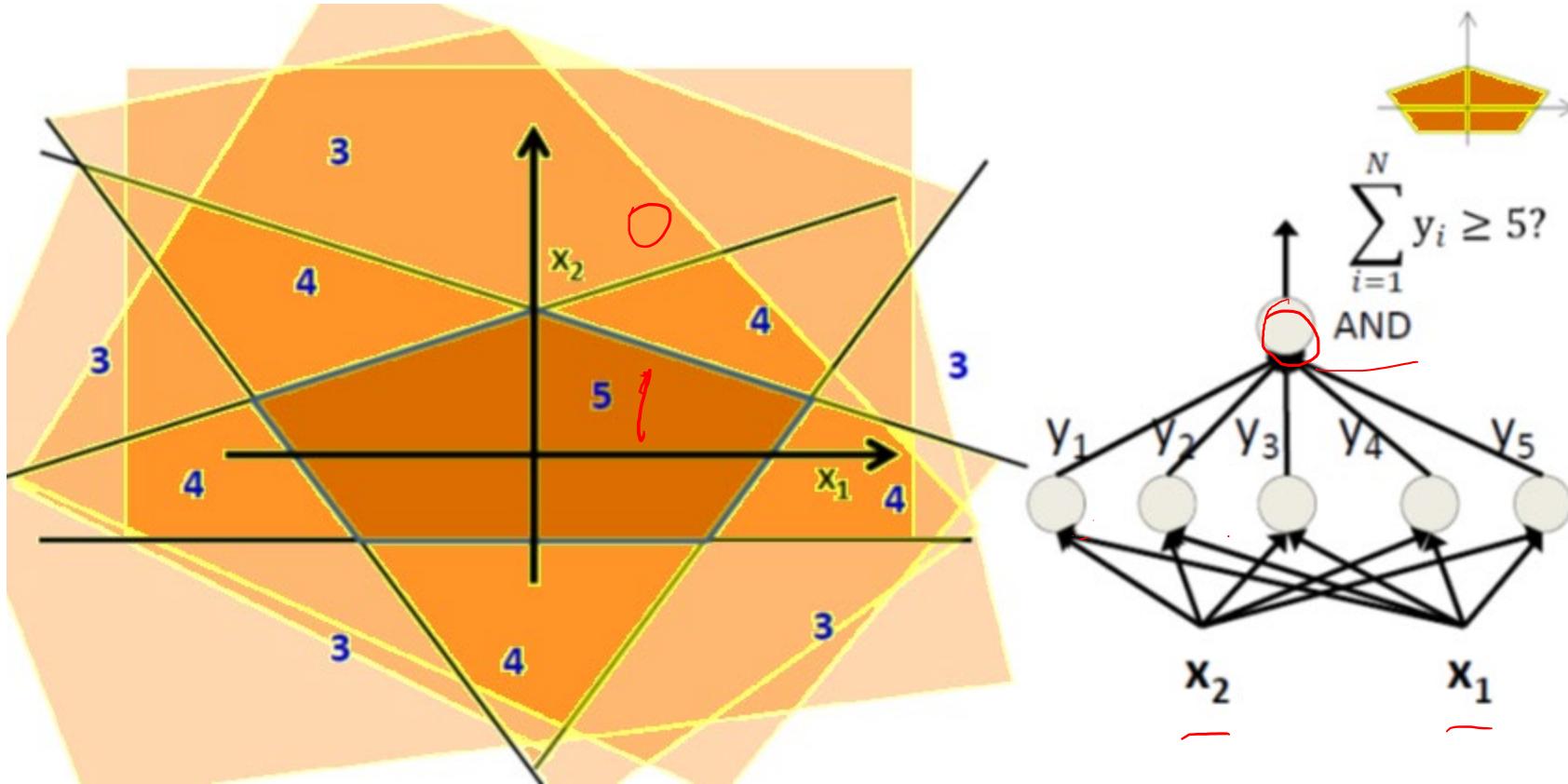
MLP



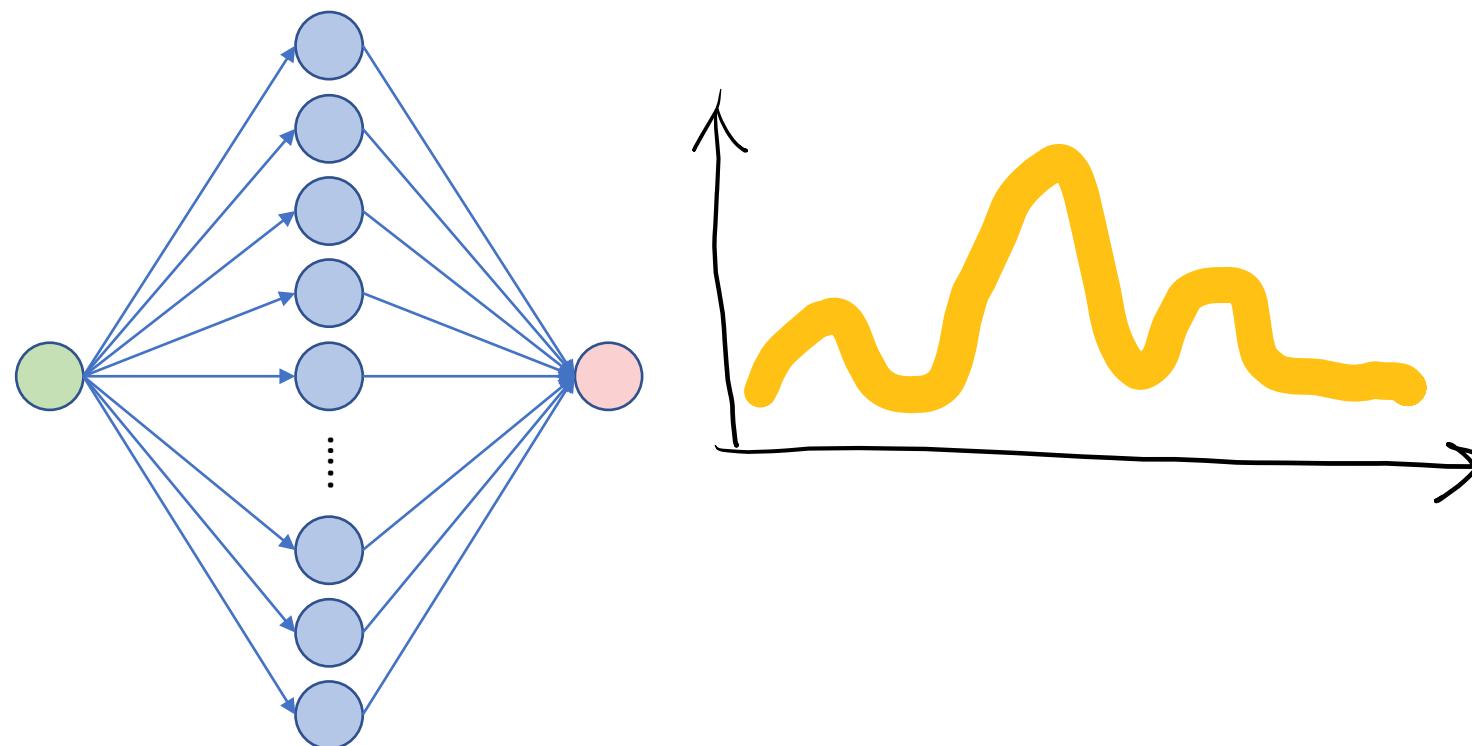
MLP



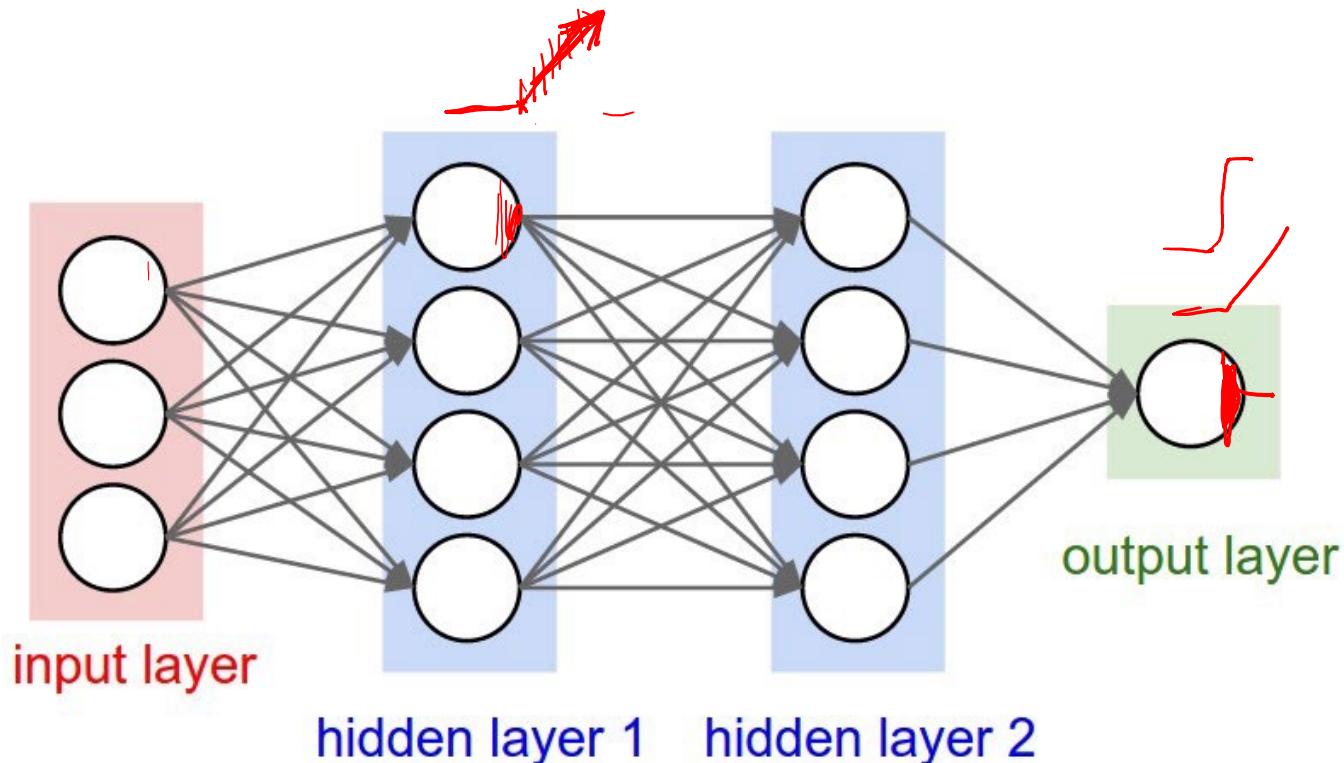
MLP



Neural networks can be used for regression tasks



Multi-layer Perceptron (Neural network)



Design Parameters

- **Architecture**
 - Number of layers
 - Number of neurons in a layer
 - Activation functions

Artificial Neural Network facts

= MLP

Can approximate any non-linear function

-> Nonlinearity comes from nonlinear activation function

Deeper layer and larger number of neurons in a layer

-> more model capacity

MLP: fully-connected neural network

Weights for every connection

FC layer
= Dense layer.



Weights = parameters ~~+~~

(cf) Hyperparameters = design parameters

Next time: How Neural Network Training Works

