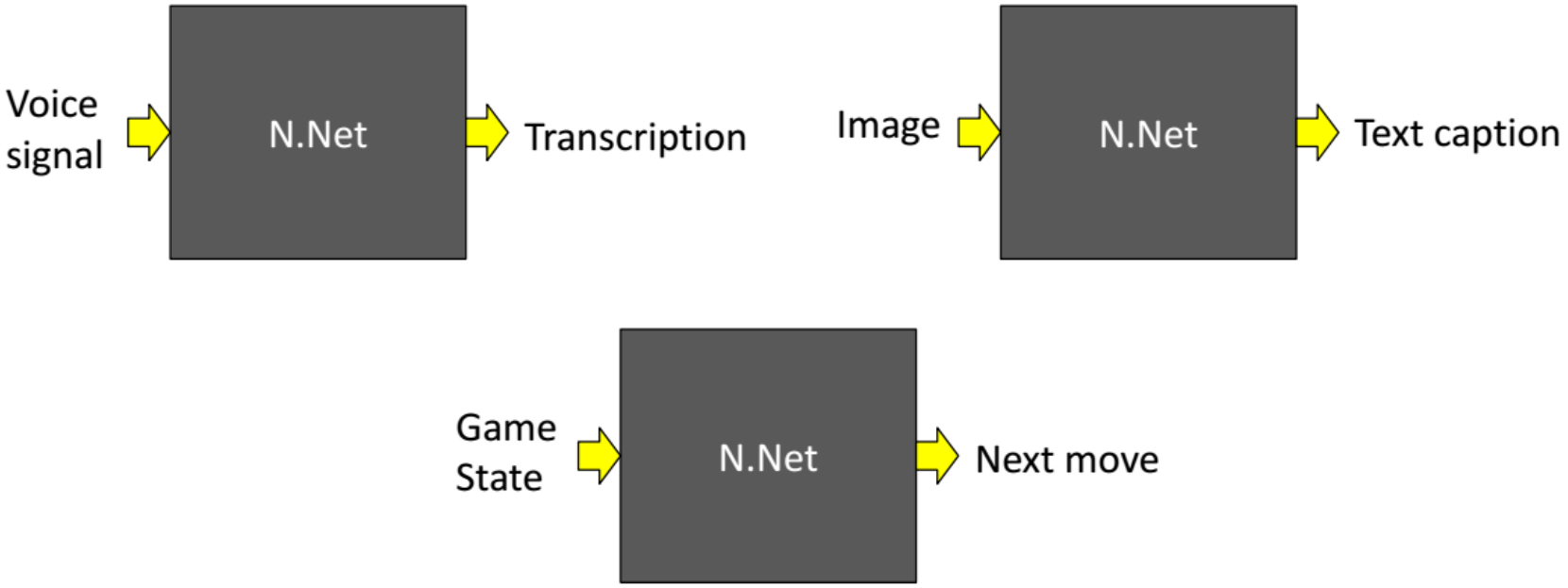
COMP417 Lecture 3

Multi-Layer Perceptrons

Dr. Hend Dawood

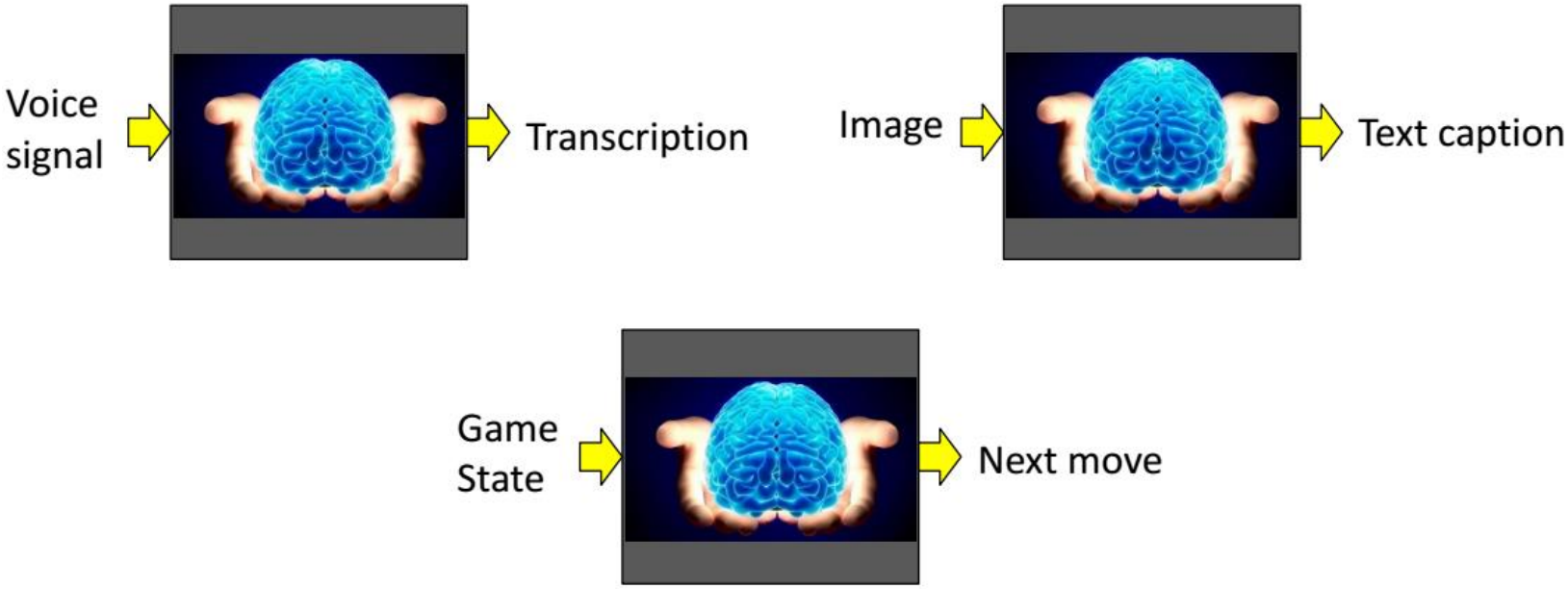
So what are neural networks?



- What are these boxes?
  - Functions that take an input and produce an output
  - What are these functions?

So what are neural networks?

The human perspective



- In a human, those functions are computed by the brain…

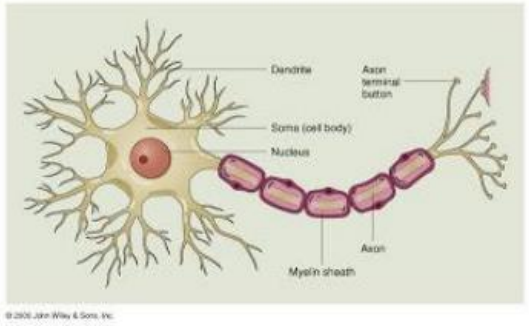
So what are neural networks?

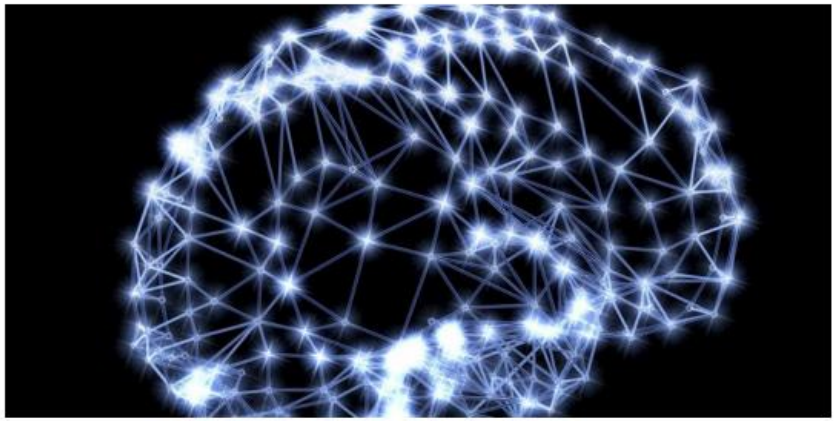
Recap : NNets and the brain

- In their basic form, NNets mimic the networked structure in the brain

So what are neural networks?

Recap : The brain

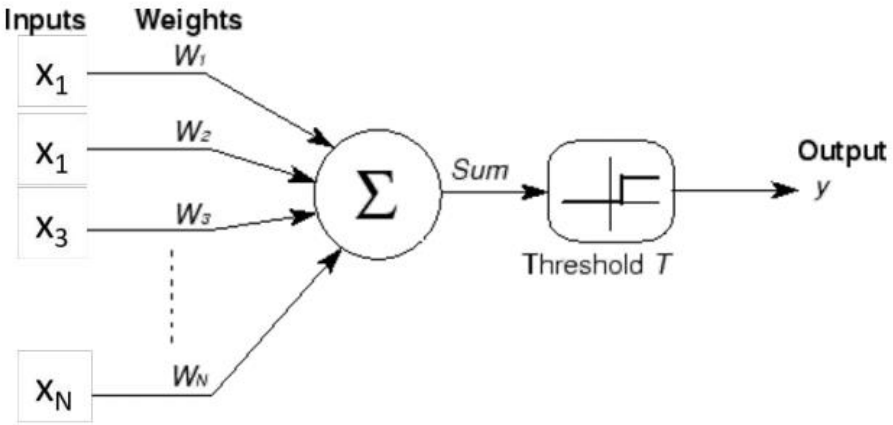


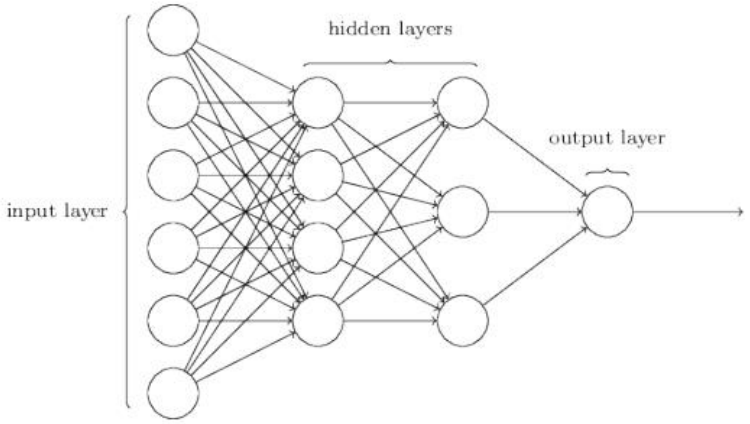


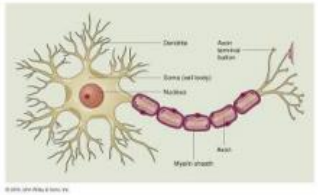
- The Brain is composed of networks of neurons

So what are neural networks?

Recap : Nnets and the brain



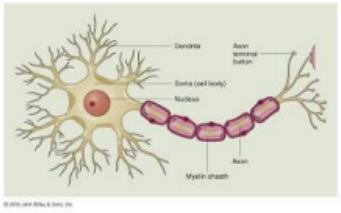


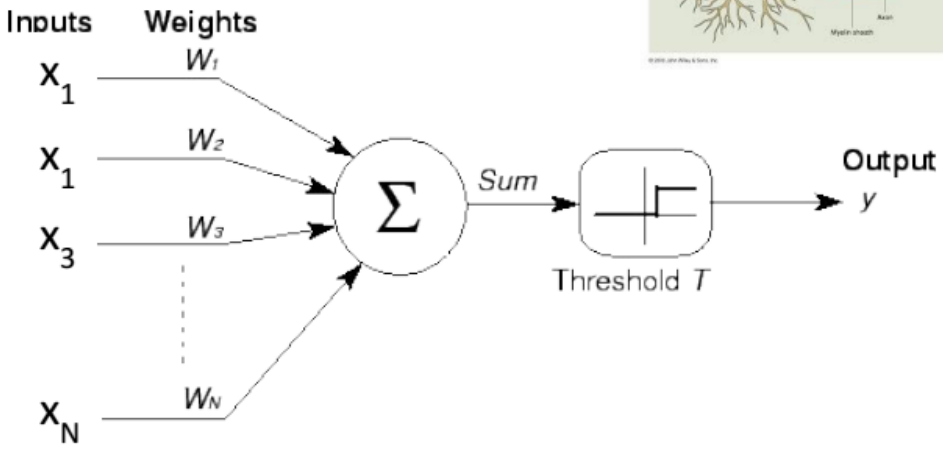


- Neural nets are composed of networks of computational models of neurons called perceptrons

So what are neural networks?

Recap: the perceptron



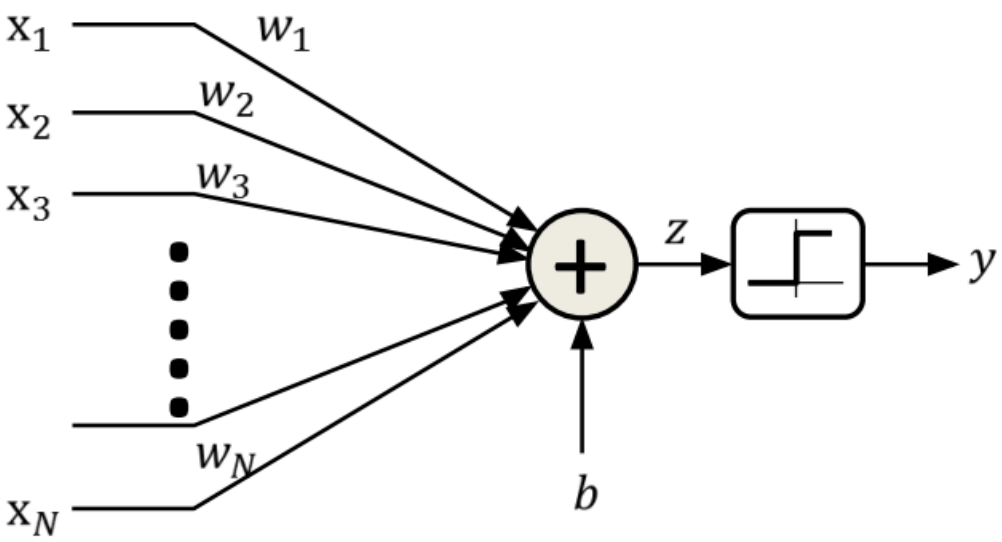


$$y = \begin{cases} 1 \; if \; \sum_i w_i x_i \geq T \\ 0 \; else \end{cases}$$

- A threshold unit
  - "Fires" if the weighted sum of inputs exceeds a threshold
  - Electrical engineers will call this a ***threshold gate***
    - A basic unit of Boolean circuits

So what are neural networks?

A better figure



Linear vs Affine?

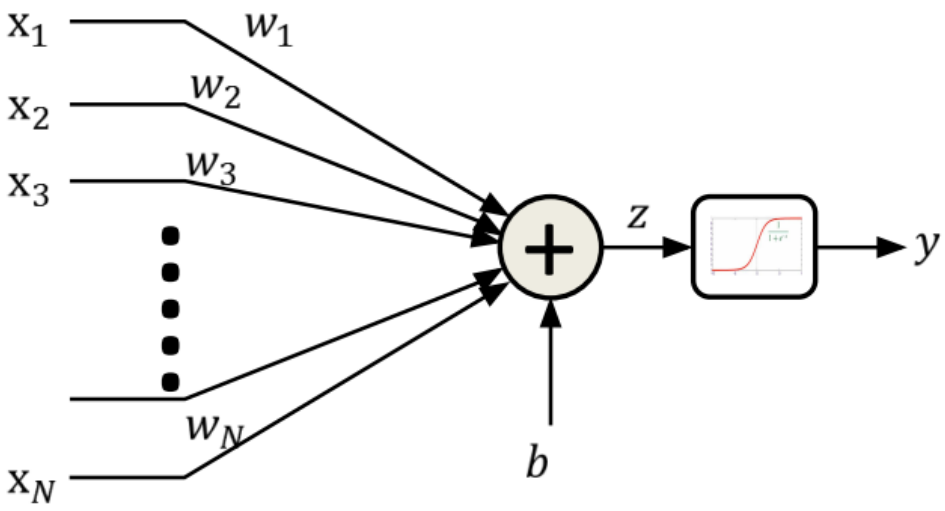
$$z = \sum_i w_i x_i + b$$

$$y = \begin{cases} 1 \; if \; z \geq 0 \\ 0 \; else \end{cases}$$

- A threshold unit
  - "Fires" if the affine function of inputs is positive
    - The bias is the negative of the threshold T in the previous slide

So what are neural networks?

The "soft" perceptron (logistic)



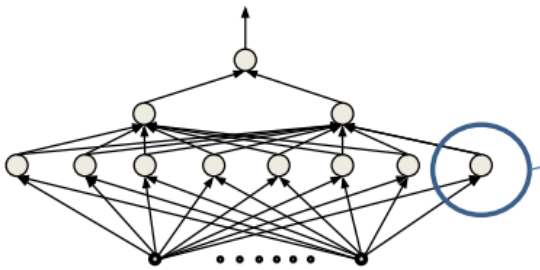
$$z = \sum_i w_i x_i + b$$

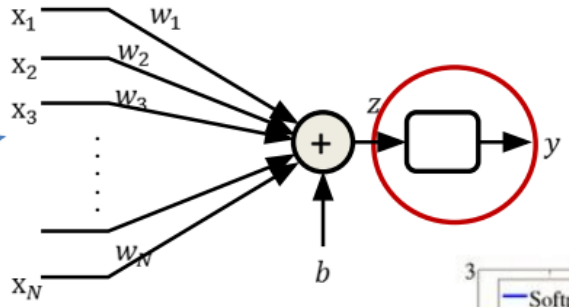
$$y = \frac{1}{1 + exp(-z)}$$

- A "squashing" function instead of a threshold at the output
  - The sigmoid "activation" replaces the threshold
    - **Activation:** The function that acts on the weighted combination of inputs (and bias)

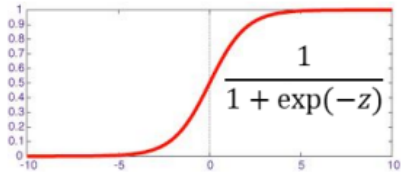
So what are neural networks?

Other "activations"

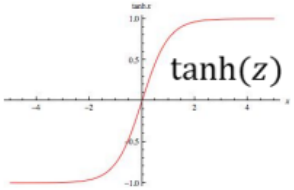


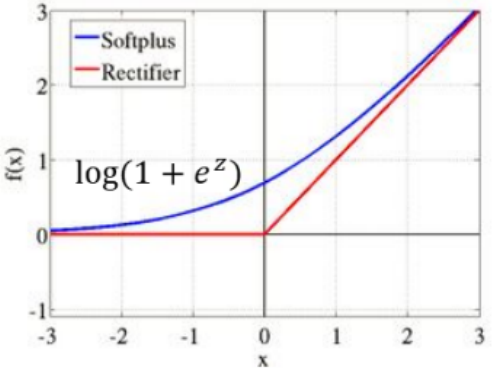


sigmoid



tanh

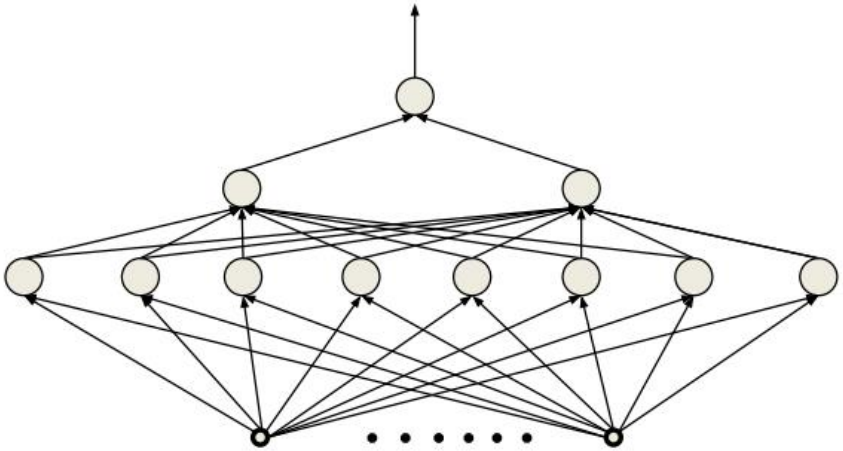


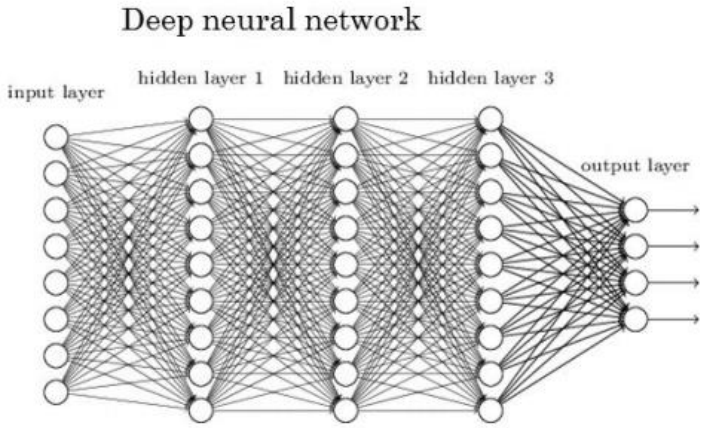


- Does not always have to be a squashing function
  - We will hear more about activations later
- We will continue to assume a "threshold" activation in this lecture

So what are neural networks?

The *multi-layer* perceptron



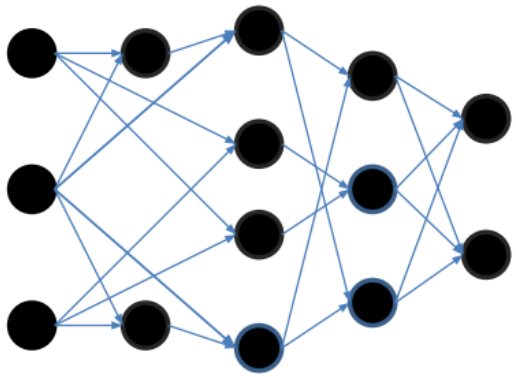


- A network of perceptrons
  - Perceptrons "feed" other perceptrons
  - We give you the "formal" definition of a layer later

So what are neural networks?

What is a layer?

- A "layer" is the set of neurons that are all at the same depth with respect to the input (sink)
  - "Depth" of a layer – the depth of the neurons in the layer w.r.t. input



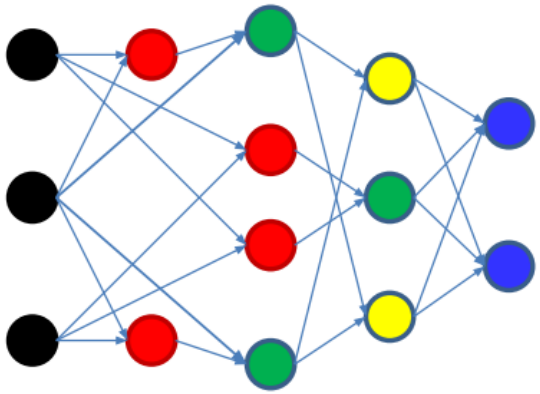
Input:
Layer 1:
Layer 2:
Layer 3:
Layer 4:

- "Deep" □ At least 3 layers
  - Output layer depth is at least 3

So what are neural networks?

What is a layer?

- A "layer" is the set of neurons that are all at the same depth with respect to the input (sink)
  - "Depth" of a layer – the depth of the neurons in the layer w.r.t. input

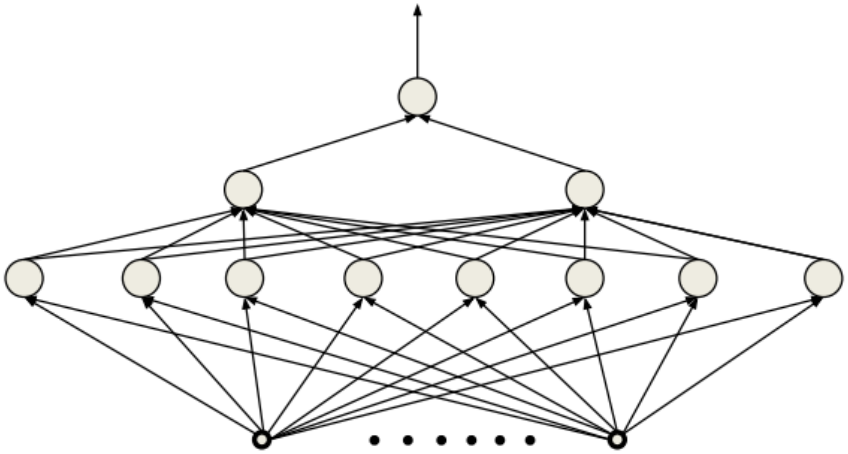


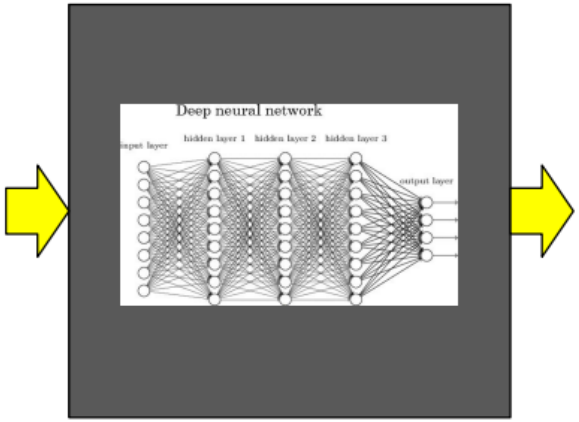
Input: Black
Layer 1: Red
Layer 2: Green
Layer 3: Yellow
Layer 4: Blue

- "Deep" □ At least 3 layers
  - Output layer depth is at least 3

So what are neural networks?

The multi-layer perceptron



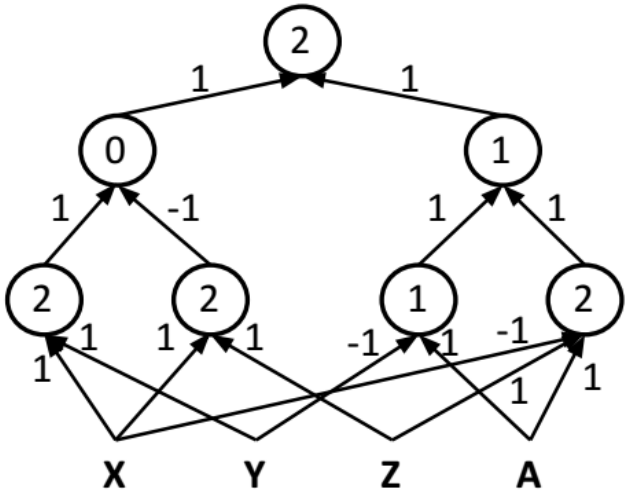


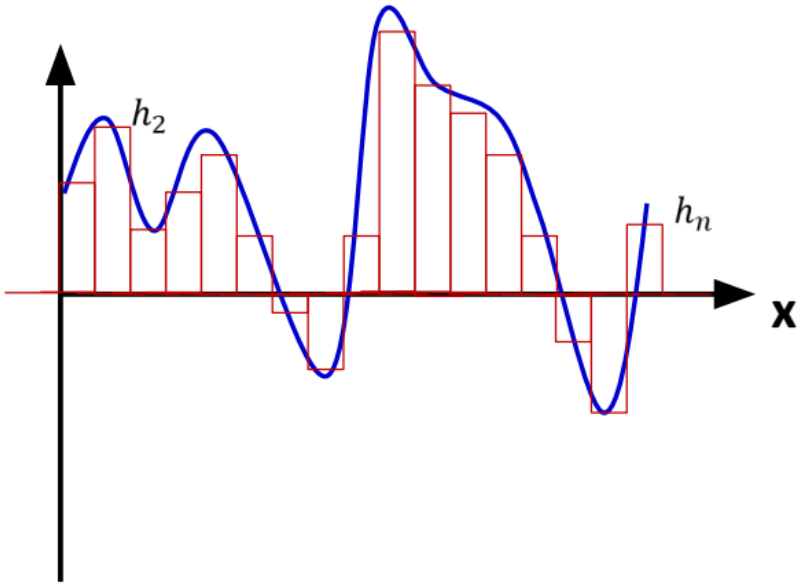
- Inputs are real or Boolean stimuli
- Outputs are real or Boolean values
  - Can have multiple outputs for a single input
- **What can this network compute?**
  - **What kinds of input/output relationships can it model?**

So what are neural networks?

MLPs approximate functions

$$((A\&\bar{X}\&Z)|(A\&\bar{Y}))\&((X\ \&\ Y)|\overline{(X\&Z)})$$

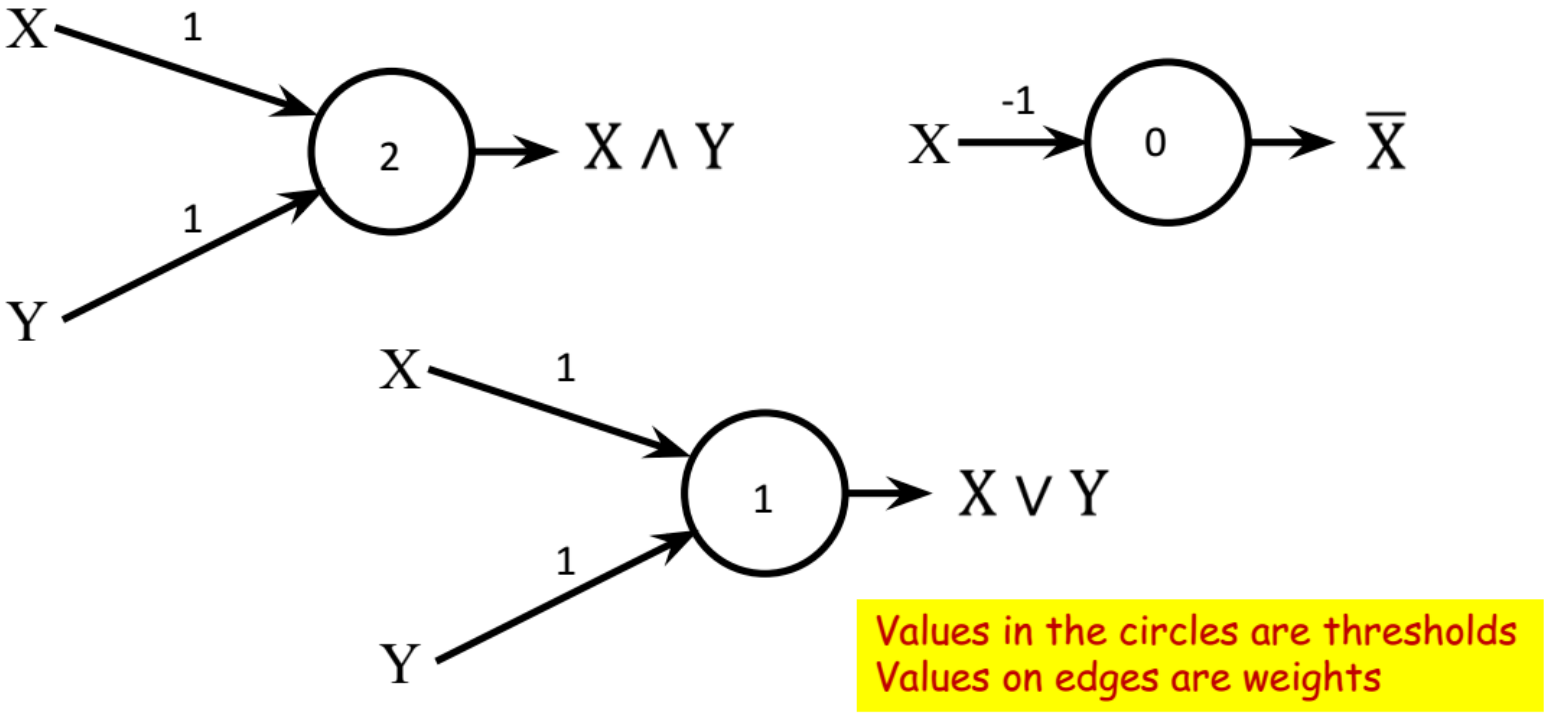




- MLPs can compose Boolean functions
- MLPs can compose real-valued functions
- What are the limitations?

So what are neural networks?

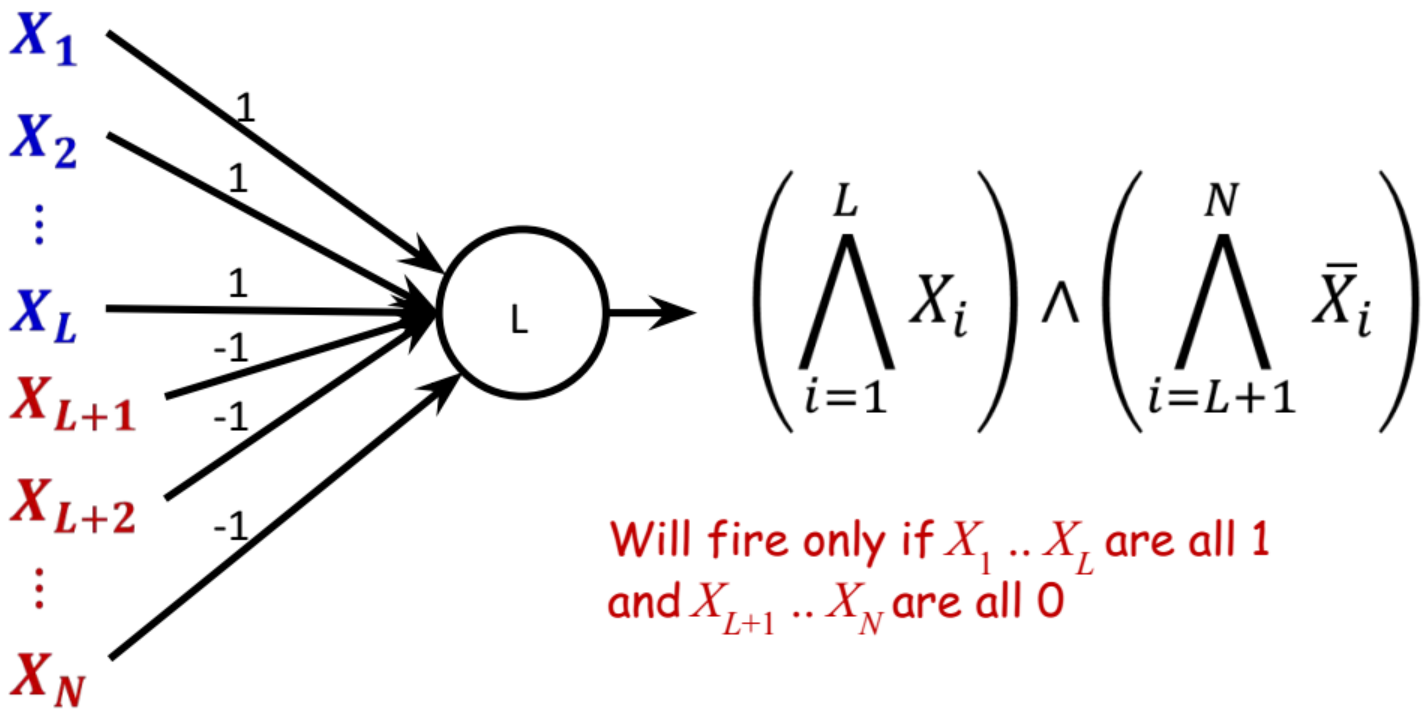
The perceptron as a Boolean gate



- A perceptron can model any simple binary Boolean gate

So what are neural networks?

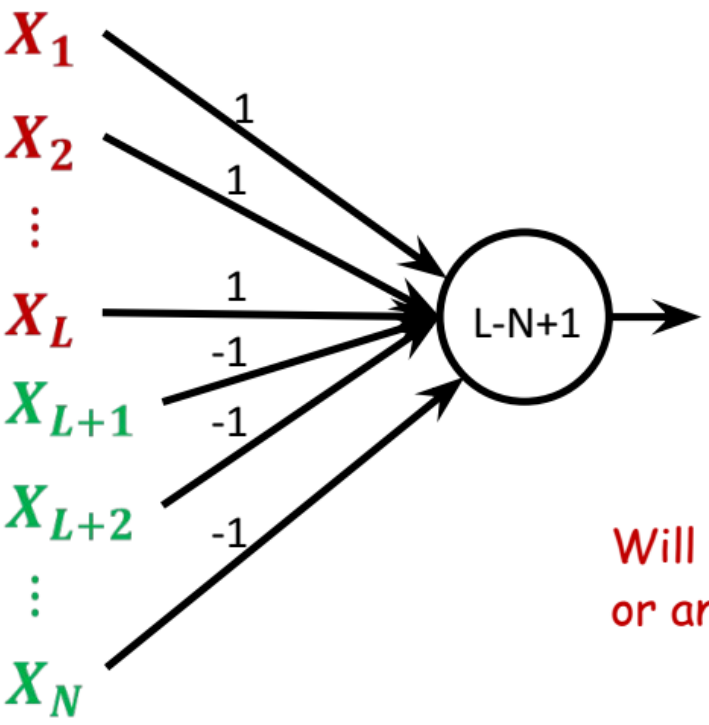
Perceptron as a Boolean gate



- The universal AND gate
  - AND any number of inputs
    - Any subset of who may be negated

So what are neural networks?

Perceptron as a Boolean gate



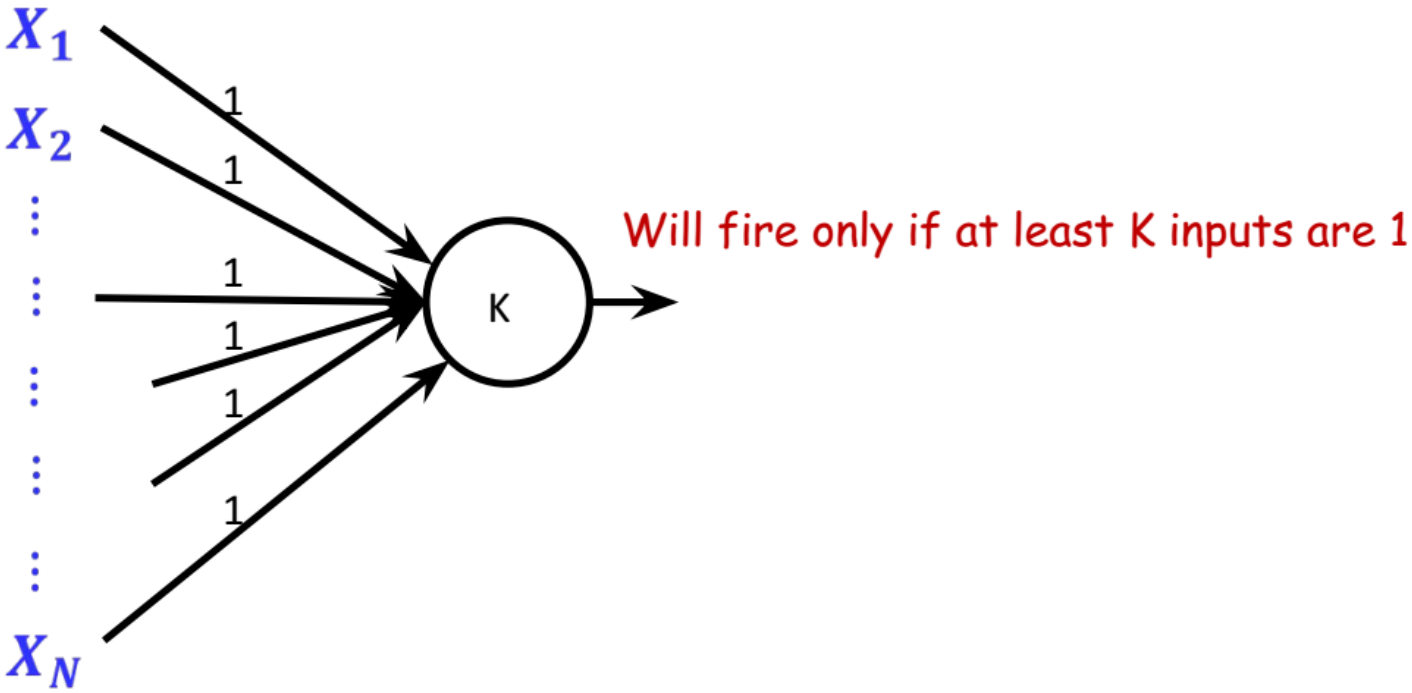
$$\left(\bigvee_{i=1}^{L} X_i\right) \vee \left(\bigvee_{i=L+1}^{N} \bar{X}_i\right)$$

Will fire only if any of $X_1 .. X_L$ are 1
or any of $X_{L+1} .. X_N$ are 0

- The universal OR gate
  - OR any number of inputs
    - Any subset of who may be negated

So what are neural networks?

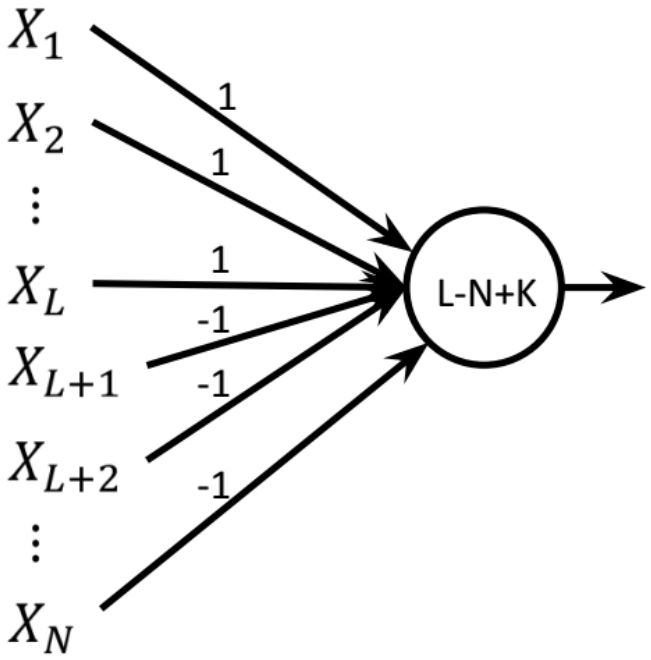
Perceptron as a Boolean Gate



- Generalized *majority* gate
  - Fire if at least *K* inputs are of the desired polarity

So what are neural networks?

Perceptron as a Boolean Gate

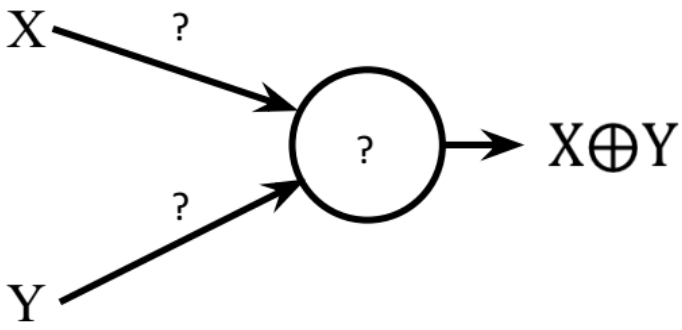


Will fire only if the total number of of X_1 .. X_L that are 1 and X_{L+1} .. X_N that are 0 is at least K

- Generalized *majority* gate
  - Fire if at least *K* inputs are of the desired polarity

So what are neural networks?

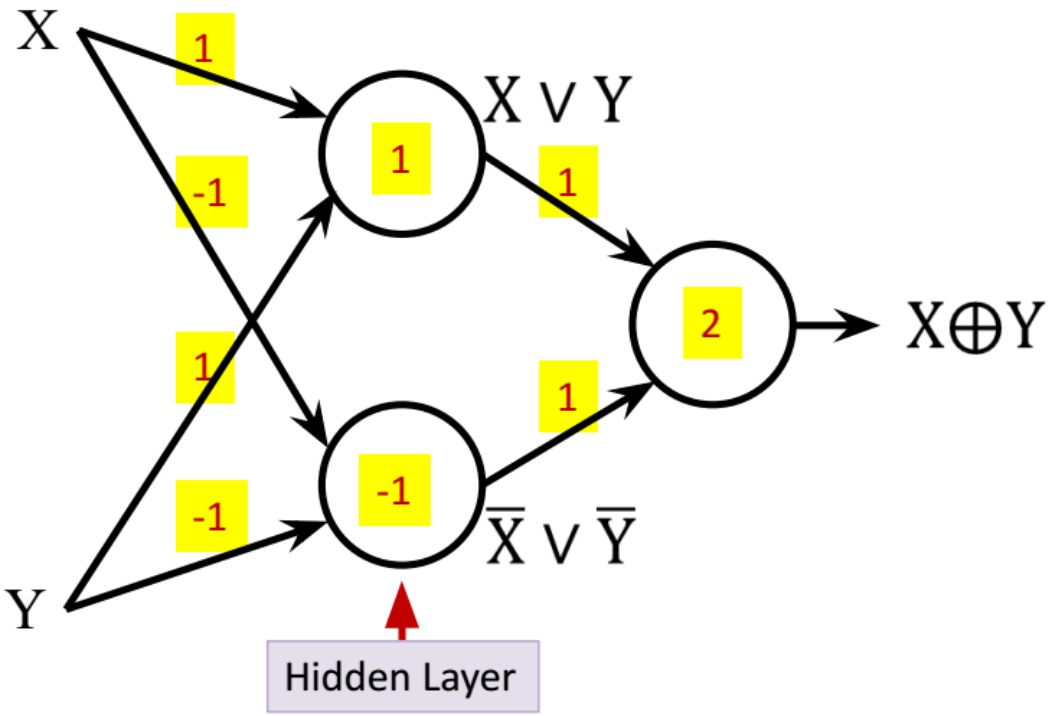
The perceptron is not enough



- Cannot compute an XOR

So what are neural networks?

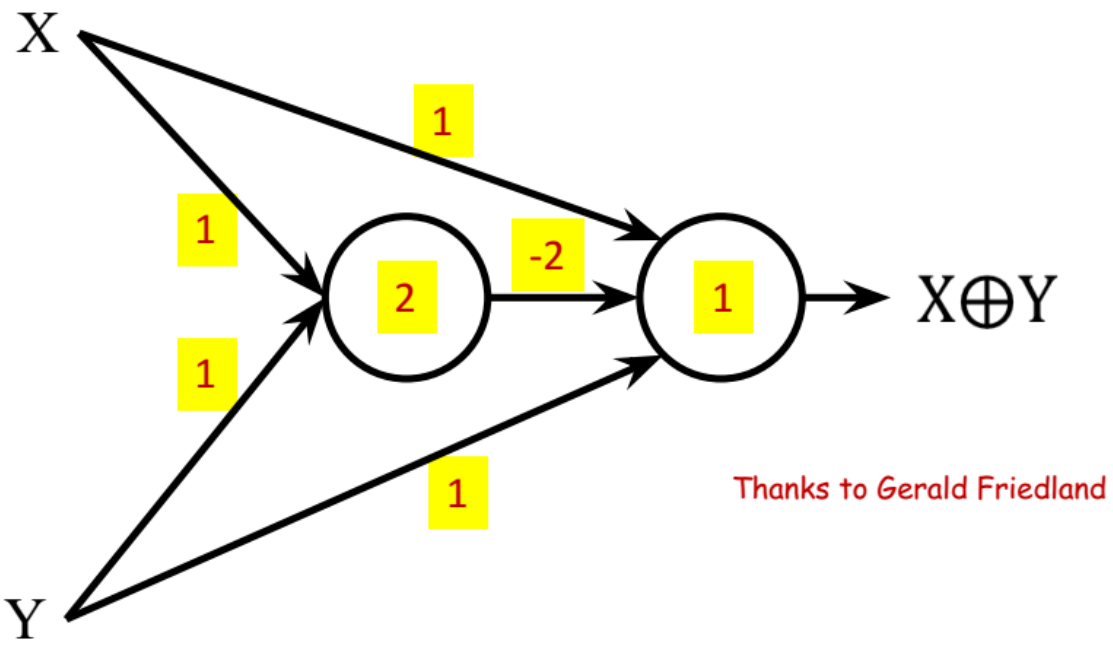
Multi-layer perceptron



- MLPs can compute the XOR

So what are neural networks?

Multi-layer perceptron XOR

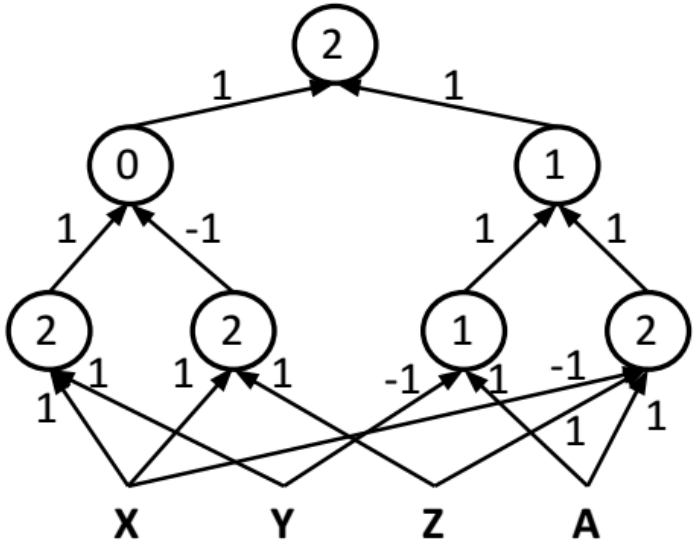


- With 2 neurons
  - 5 weights and two thresholds

So what are neural networks?

Multi-layer perceptron

$$((A\&\bar{X}\&Z)|(A\&\bar{Y}))\&((X\ \&\ Y)|\overline{(X\&Z)})$$

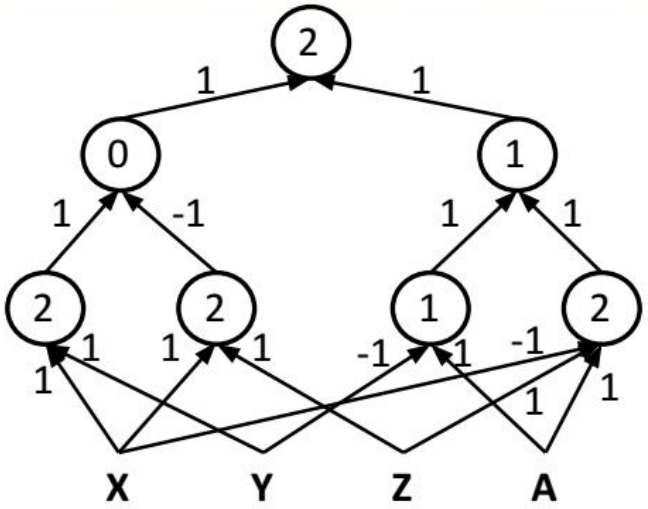


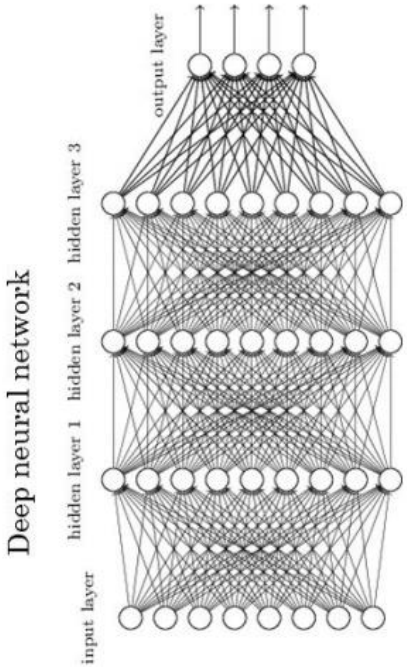
- MLPs can compute more complex Boolean functions
- MLPs can compute *any* Boolean function
  - Since they can emulate individual gates
- **MLPs are *universal Boolean functions***

So what are neural networks?

MLP as Boolean Functions

$$((A\&\bar{X}\&Z)|(A\&\bar{Y}))\&((X\ \&\ Y)|\overline{(X\&Z)})$$





- MLPs are universal Boolean functions
  - Any function over any number of inputs and any number of outputs
- But how many "layers" will they need?

So what are neural networks?

How many layers for a Boolean MLP?

Truth table shows *all* input combinations for which output is 1

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

- ***A Boolean function is just a truth table***

So what are neural networks?

How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$

- Expressed in disjunctive normal form

So what are neural networks?

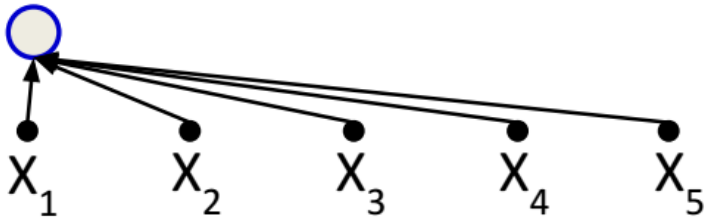
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

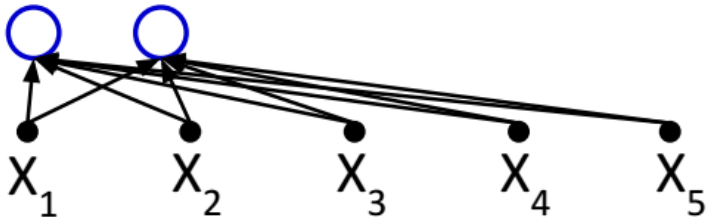
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

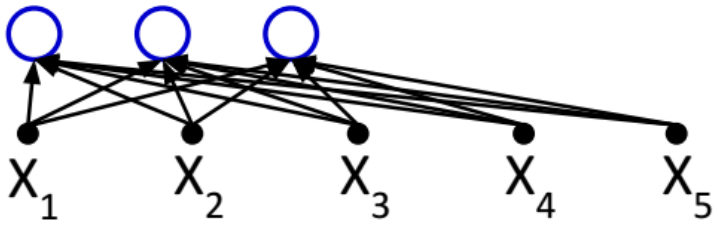
How many layers for a Boolean MLP?

Truth table shows *all* input combinations for which output is 1

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3X_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

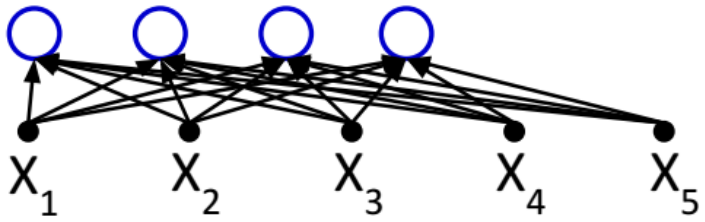
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

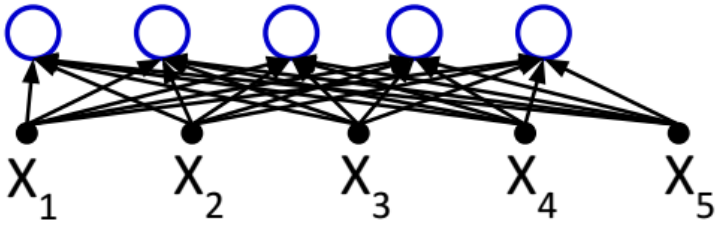
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

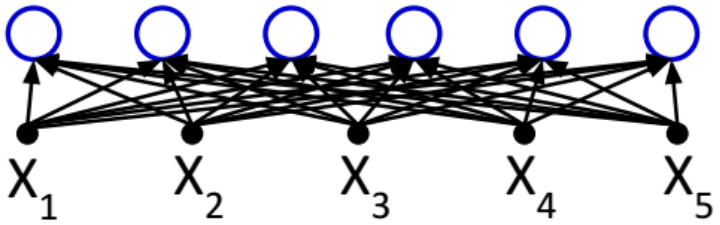
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

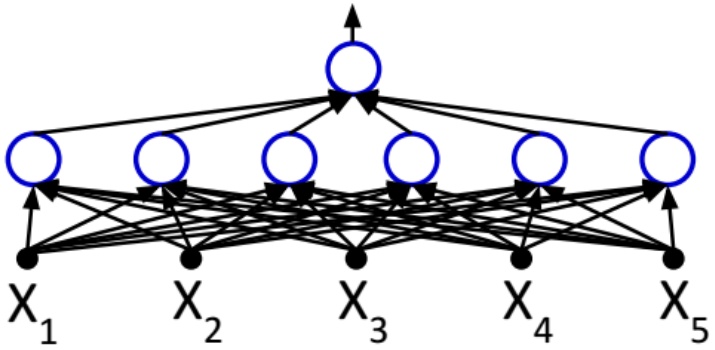
How many layers for a Boolean MLP?

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

Truth table shows *all* input combinations for which output is 1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- Expressed in disjunctive normal form

So what are neural networks?

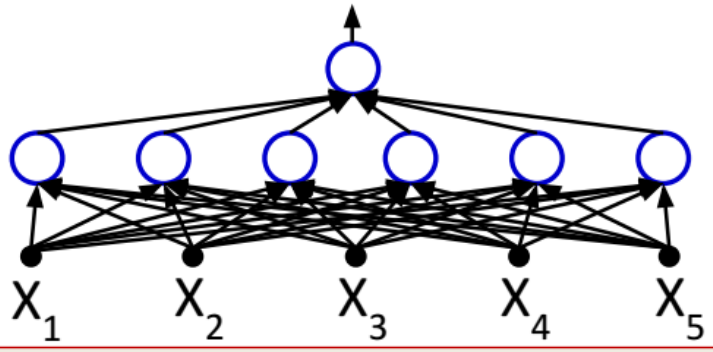
How many layers for a Boolean MLP?

Truth table shows *all* input combinations for which output is 1

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$



- *Any truth table can be expressed in this manner!*
- **A one-hidden-layer MLP is a Universal Boolean Function**

So what are neural networks?

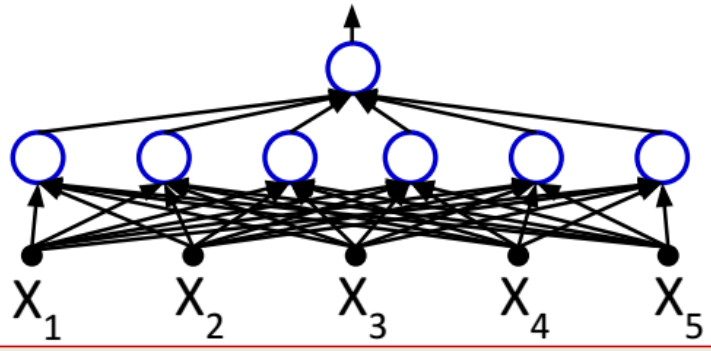
How many layers for a Boolean MLP?

Truth table shows *all* input combinations for which output is 1

Truth Table

X_1	X_2	X_3	X_4	X_5	Y
0	0	1	1	0	1
0	1	0	1	1	1
0	1	1	0	0	1
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	1	1

$$Y = \bar{X}_1\bar{X}_2X_3X_4\bar{X}_5 + \bar{X}_1X_2\bar{X}_3X_4X_5 + \bar{X}_1X_2X_3\bar{X}_4\bar{X}_5 + X_1\bar{X}_2\bar{X}_3\bar{X}_4X_5 + X_1\bar{X}_2X_3X_4X_5 + X_1X_2\bar{X}_3\bar{X}_4X_5$$

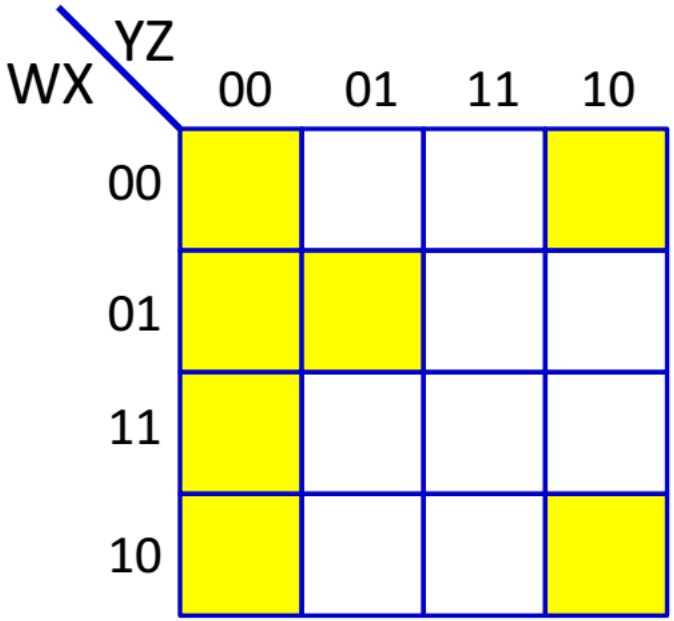


- *Any truth table can be expressed in this manner!*
- **A one-hidden-layer MLP is a Universal Boolean Function**

But what is the largest number of perceptrons required in the single hidden layer for an N-input-variable function?

So what are neural networks?

Reducing a Boolean Function



This is a "Karnaugh Map"

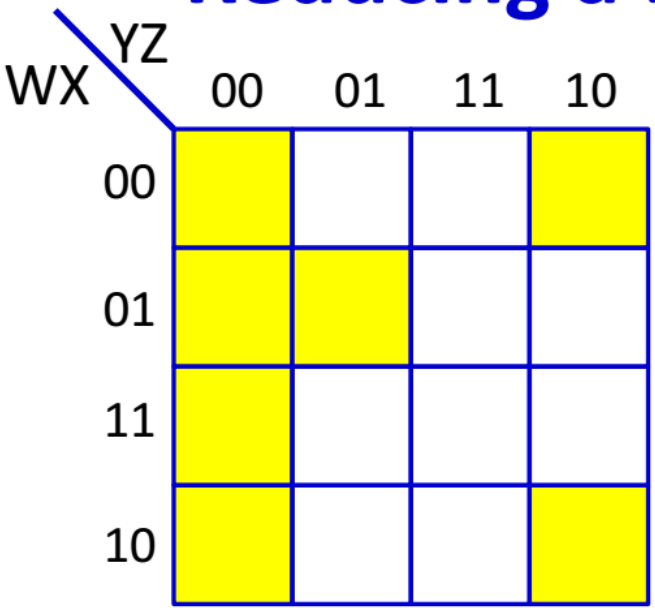
It represents a truth table as a grid
Filled boxes represent input combinations for which output is 1; blank boxes have output 0

Adjacent boxes can be "grouped" to reduce the complexity of the DNF formula for the table

- DNF form:
  - Find groups
  - Express as reduced DNF

So what are neural networks?

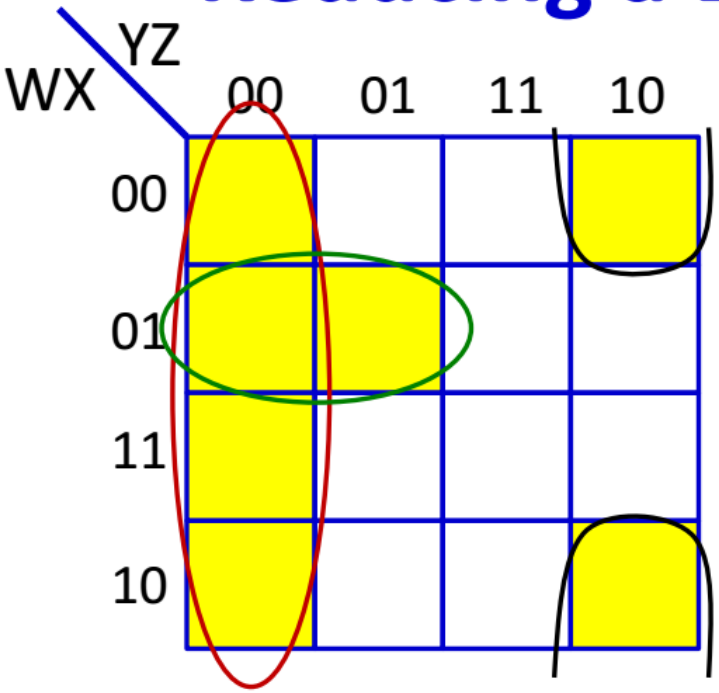
Reducing a Boolean Function



Basic DNF formula will require 7 terms

So what are neural networks?

Reducing a Boolean Function

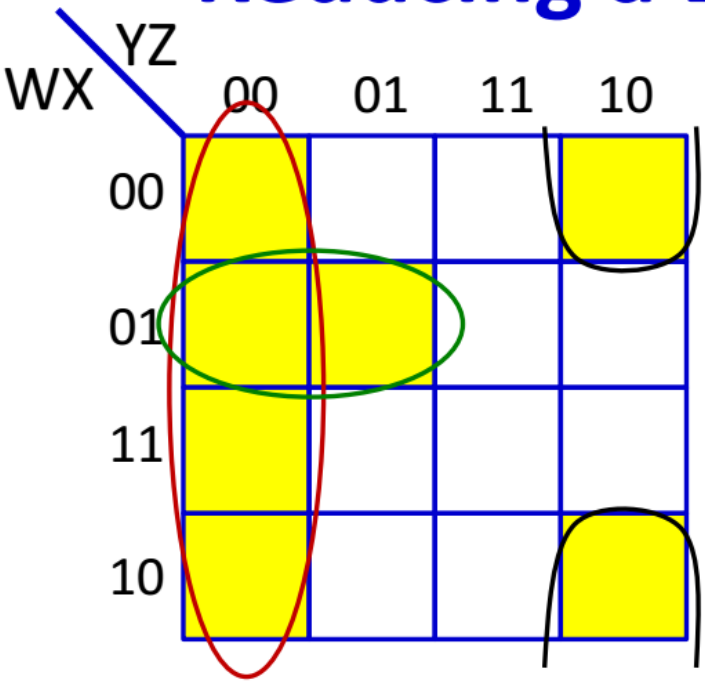


$$O = \bar{Y}\bar{Z} + \bar{W}X\bar{Y} + \bar{X}Y\bar{Z}$$

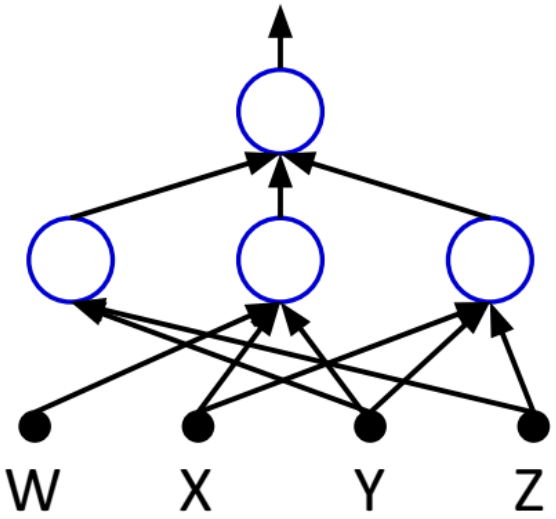
- *Reduced* DNF form:
  - Find groups
  - Express as reduced DNF

So what are neural networks?

Reducing a Boolean Function



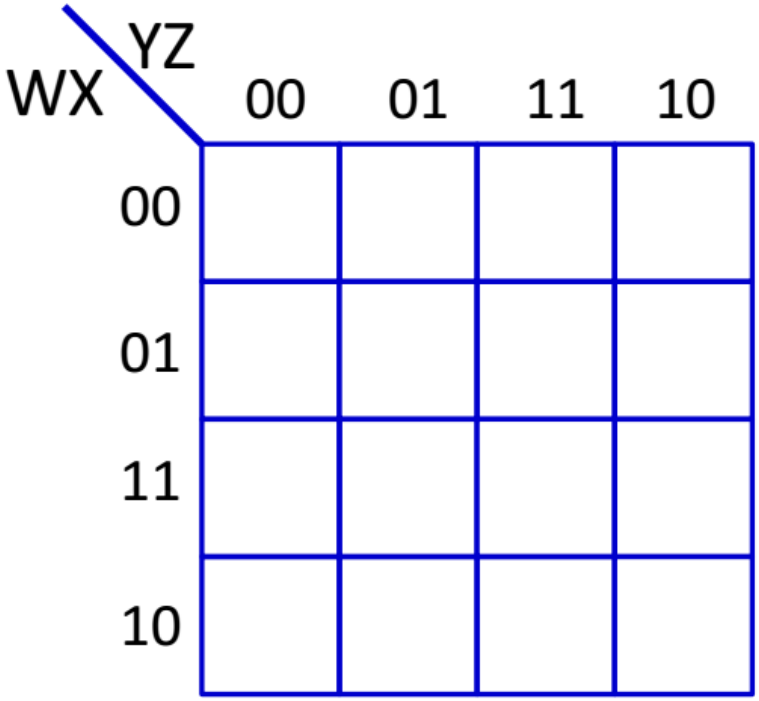
$$O = \bar{Y}\bar{Z} + \bar{W}X\bar{Y} + \bar{X}Y\bar{Z}$$



- *Reduced* DNF form:
  - Find groups
  - Express as *reduced* DNF
  - Boolean network for this function needs only 3 hidden units
    - Reduction of the DNF reduces the size of the one-hidden-layer network

So what are neural networks?

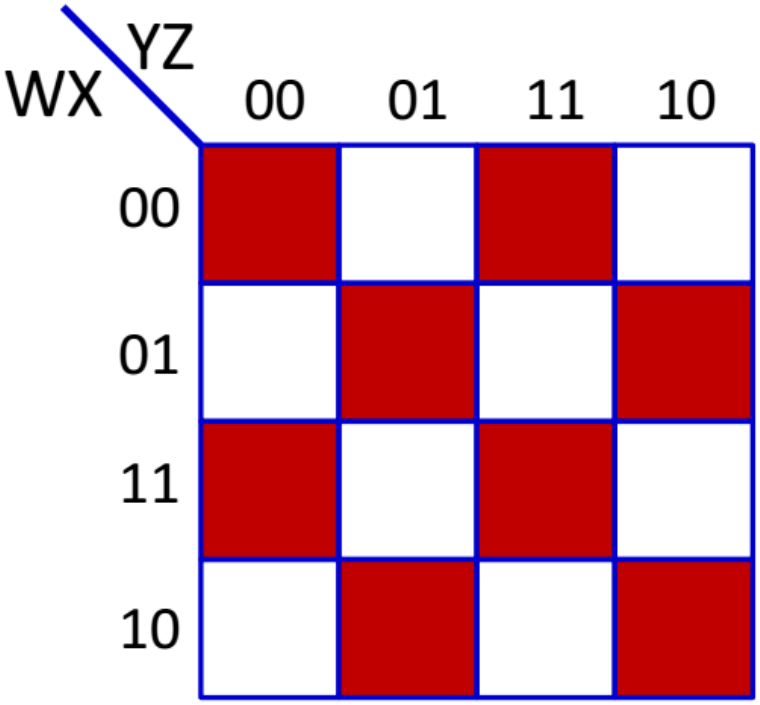
Largest irreducible DNF?



- What arrangement of ones and zeros simply cannot be reduced further?

So what are neural networks?

Largest irreducible DNF?

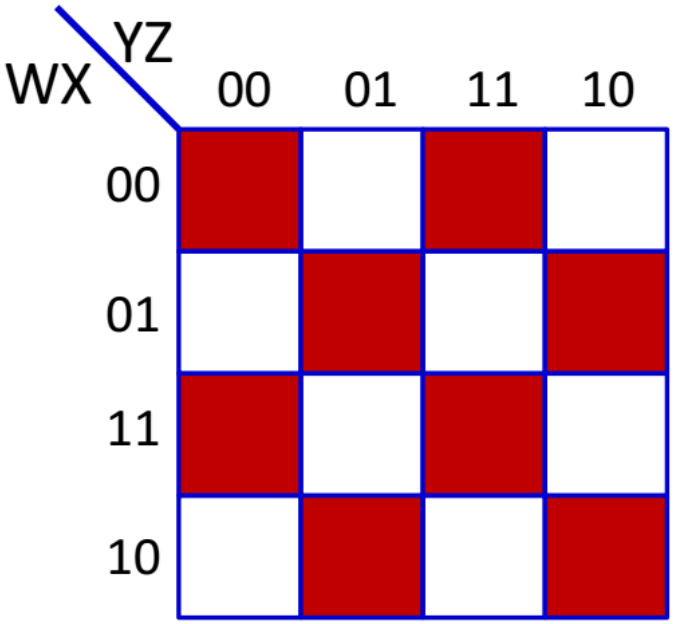


Red=0, white=1

- What arrangement of ones and zeros simply cannot be reduced further?

So what are neural networks?

Largest irreducible DNF?

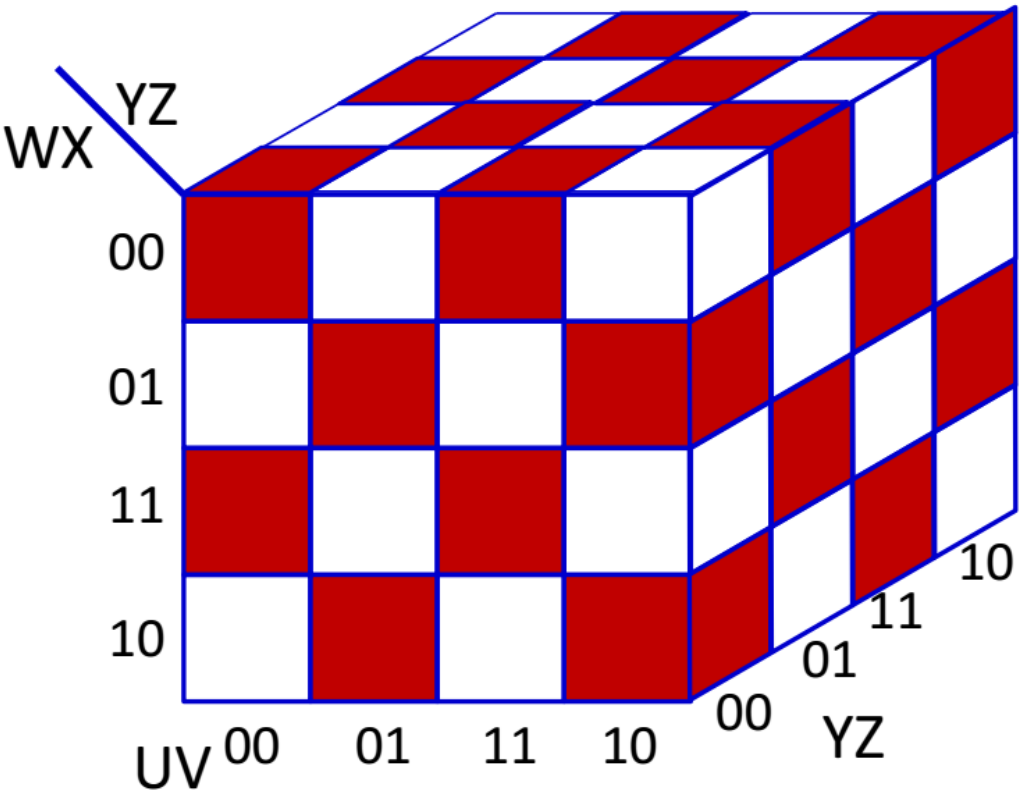


How many neurons in a DNF (one-hidden-layer) MLP for this Boolean function?

- What arrangement of ones and zeros simply cannot be reduced further?

So what are neural networks?

Width of a one-hidden-layer Boolean MLP

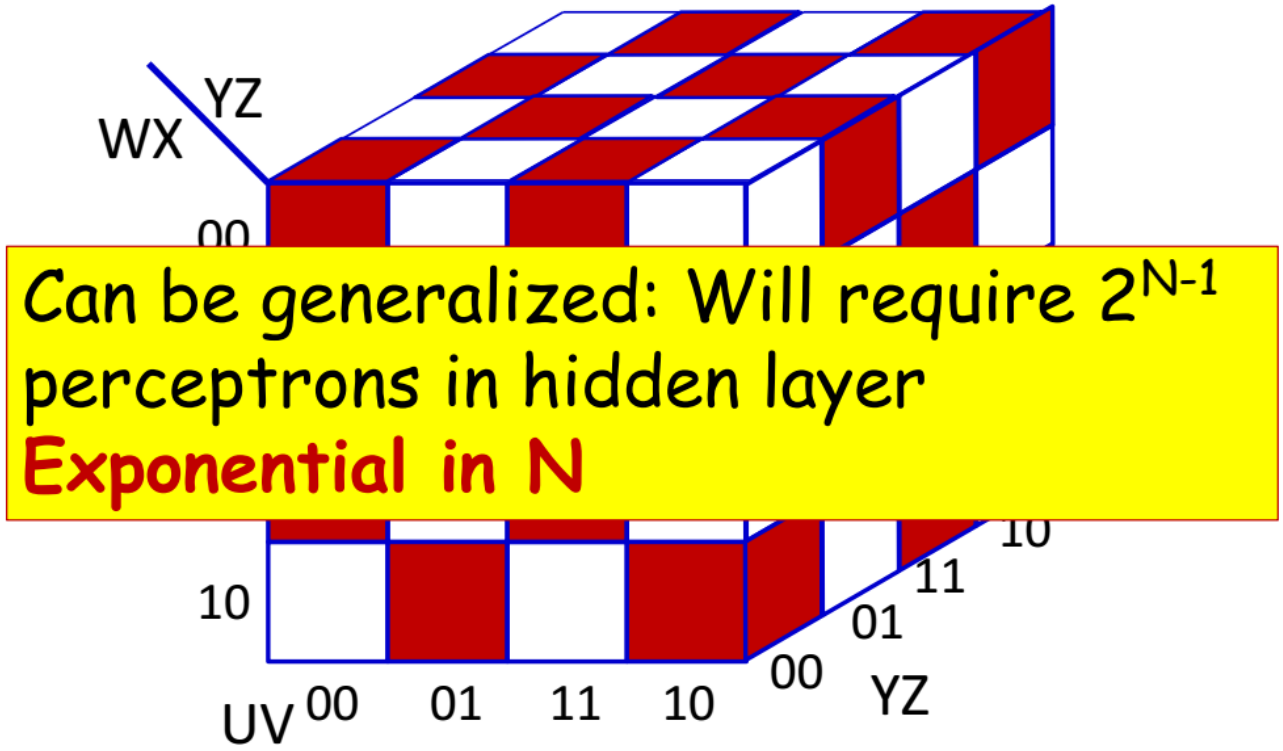


Red=0, white=1

- How many neurons in a DNF (one-hidden-layer) MLP for this Boolean function of 6 variables?

So what are neural networks?

Width of a one-hidden-layer Boolean MLP

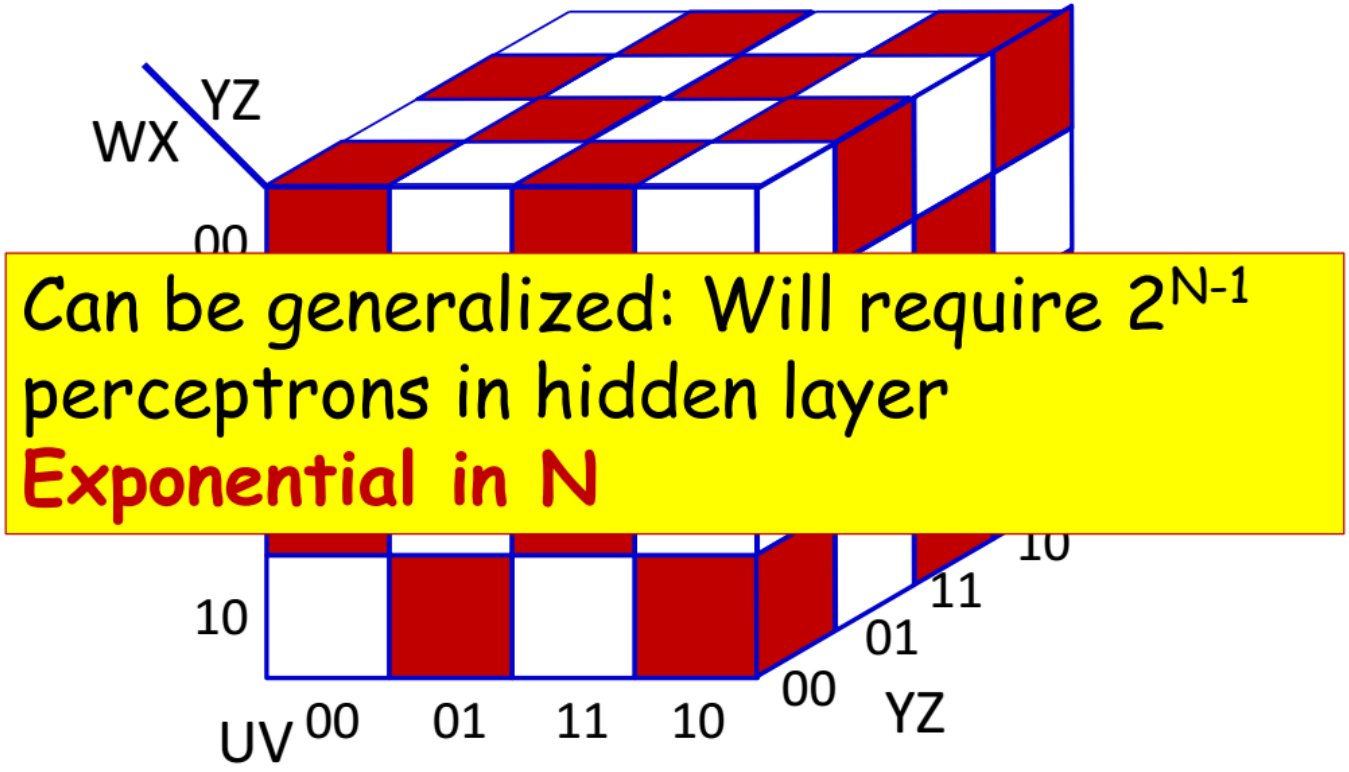


Can be generalized: Will require 2^{N-1} perceptrons in hidden layer
Exponential in N

- How many neurons in a DNF (one-hidden-layer) MLP for this Boolean function

So what are neural networks?

Width of a one-hidden-layer Boolean MLP



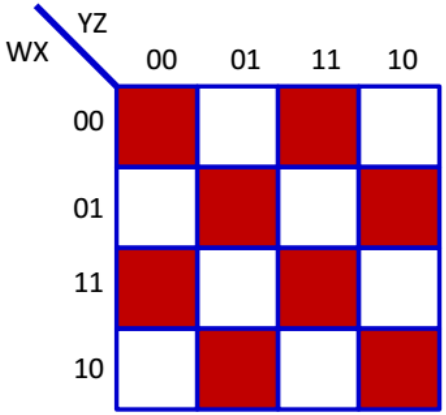
Can be generalized: Will require 2^{N-1} perceptrons in hidden layer

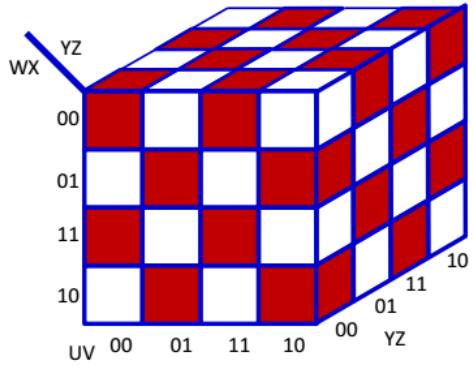
Exponential in N

How many units if we use multiple hidden layers?

So what are neural networks?

Size of a deep MLP



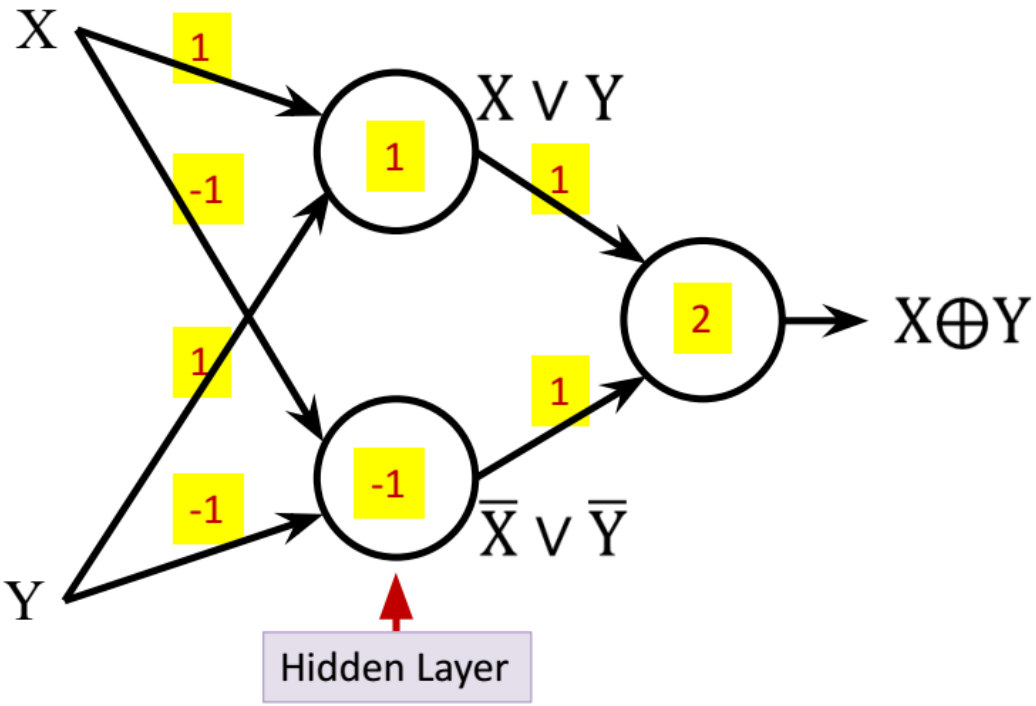


$$O = W \oplus X \oplus Y \oplus Z$$

$$O = U \oplus V \oplus W \oplus X \oplus Y \oplus Z$$

So what are neural networks?

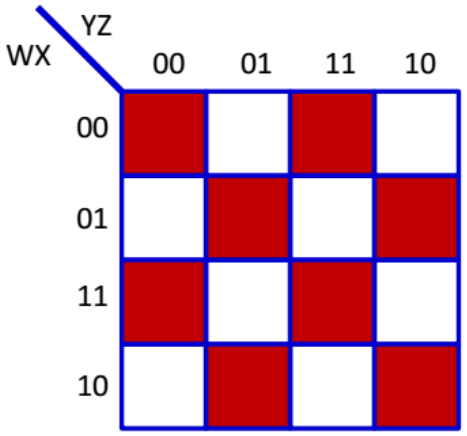
Multi-layer perceptron XOR



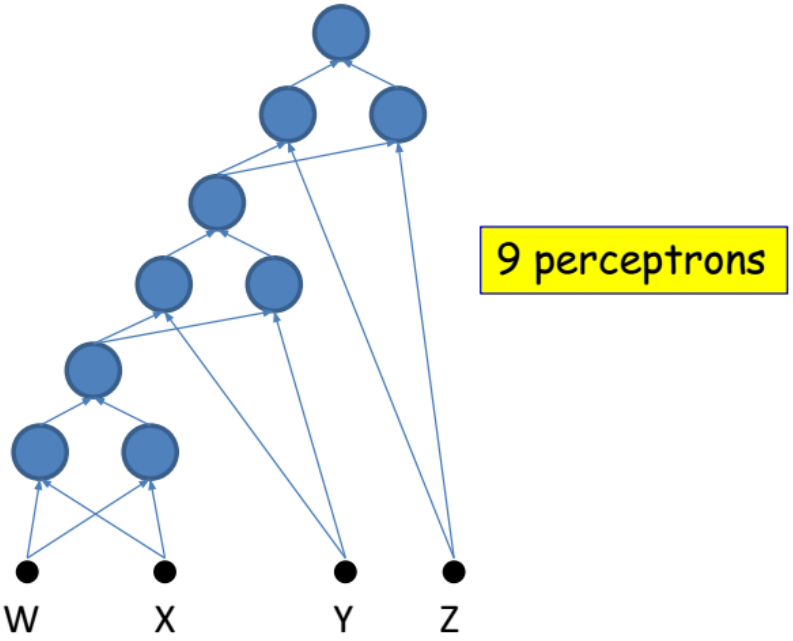
- An XOR takes three perceptrons

So what are neural networks?

Size of a deep MLP



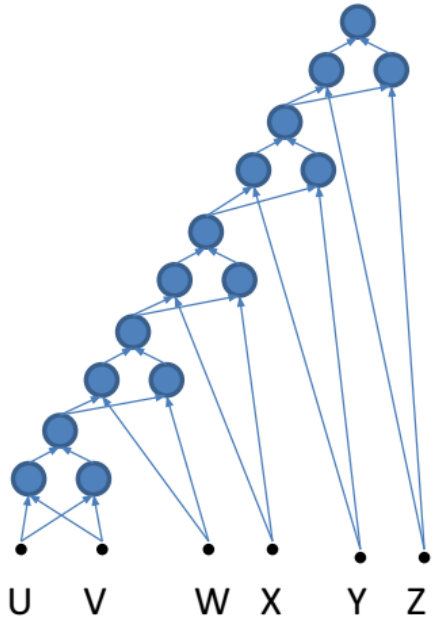
$$O = W \oplus X \oplus Y \oplus Z$$

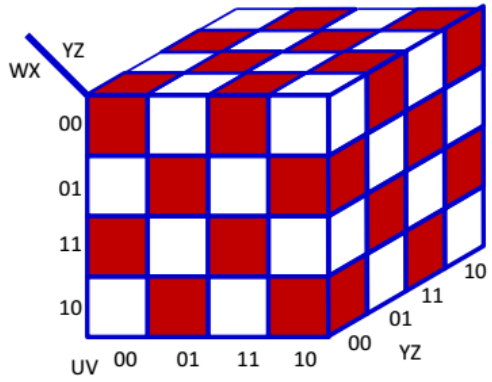


- An XOR needs 3 perceptrons
- This network will require 3x3 = 9 perceptrons

So what are neural networks?

Size of a deep MLP





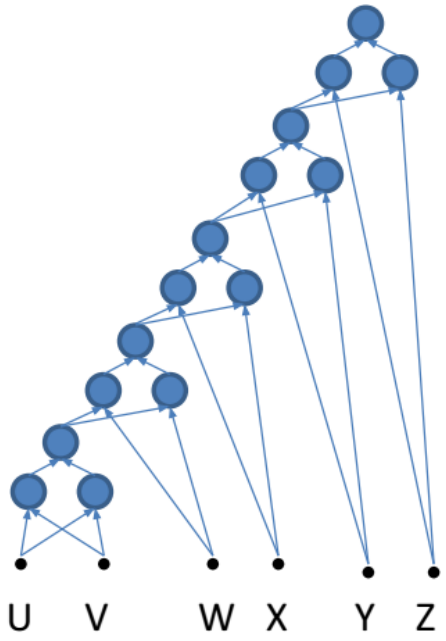
$$O = U \oplus V \oplus W \oplus X \oplus Y \oplus Z$$

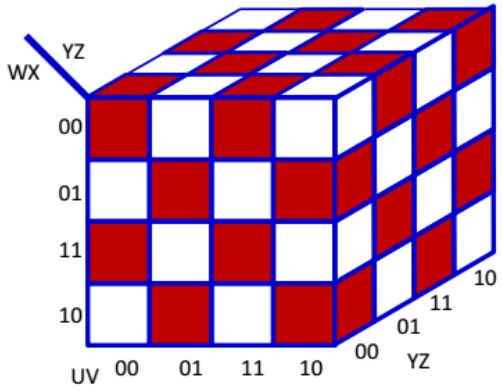
15 perceptrons

- An XOR needs 3 perceptrons
- This network will require 3x5 = 15 perceptrons

So what are neural networks?

Size of a deep MLP





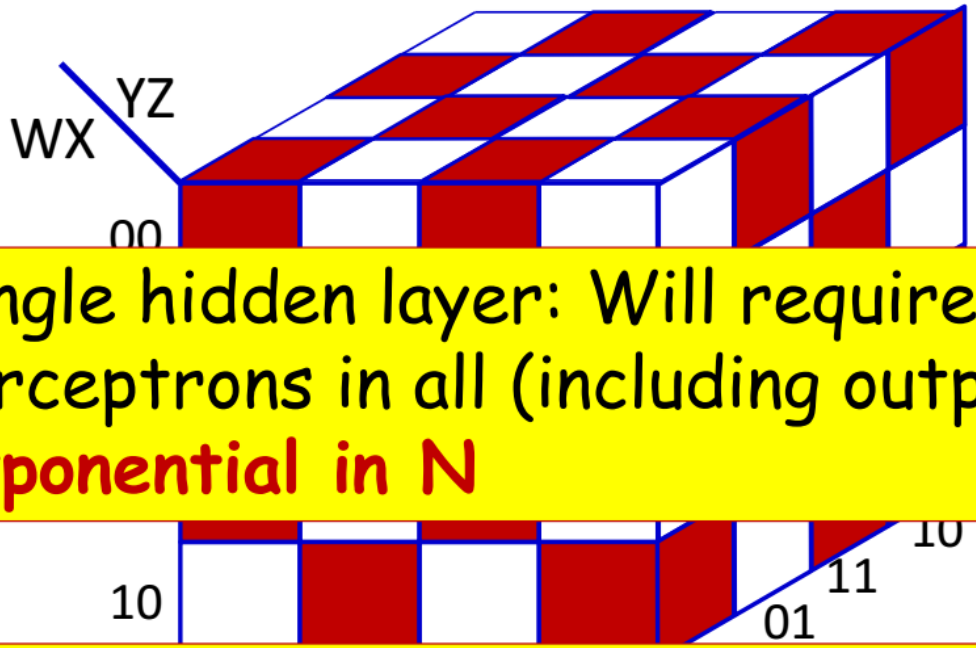
$$O = U \oplus V \oplus W \oplus X \oplus Y \oplus Z$$

More generally, the XOR of N variables will require 3(N-1) perceptrons!!

- An XOR needs 3 perceptrons
- This network will require 3x5 = 15 perceptrons

So what are neural networks?

One-hidden layer vs deep Boolean MLP

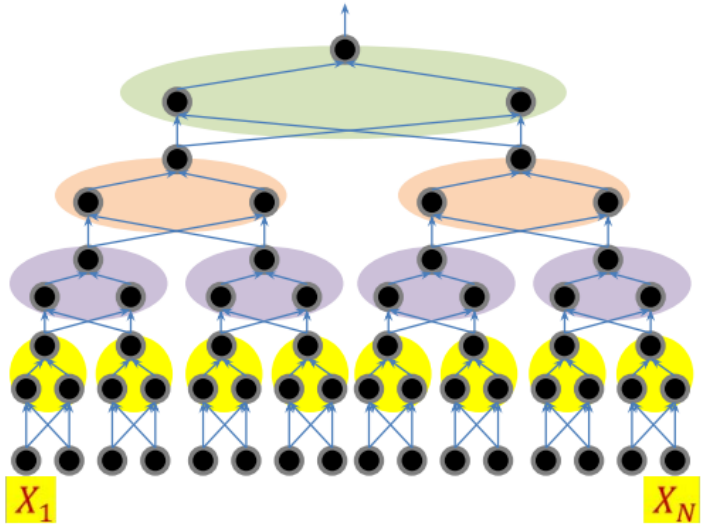


Single hidden layer: Will require $2^{N-1}+1$ perceptrons in all (including output unit)
Exponential in N

Will require 3(N-1) perceptrons in a deep network
Linear in N!!!
Can be arranged in only $2\log_2(N)$ layers

So what are neural networks?

A better representation



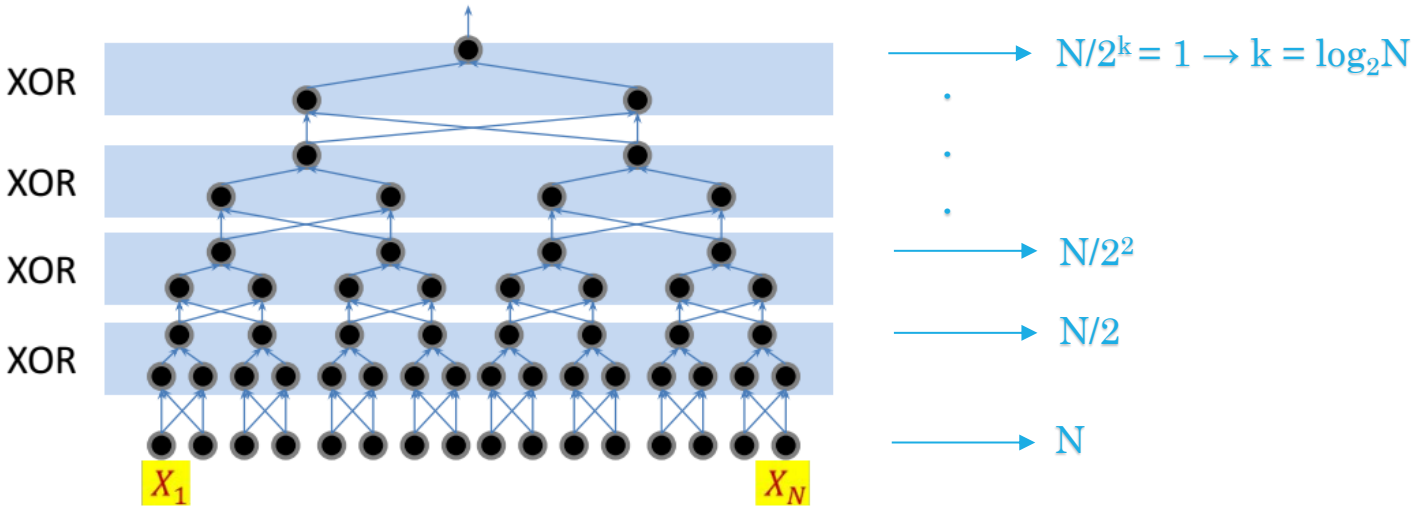
$$O = X_1 \oplus X_2 \oplus \cdots \oplus X_N$$

- Only $2\log_2 N$ layers
  - By pairing terms
  - 2 layers per XOR

$$O = ((((X_1 \oplus X_2) \oplus (X_3 \oplus X_4)) \oplus ((X_5 \oplus X_6) \oplus (X_7 \oplus X_8))) \oplus (((...$$

So what are neural networks?

A better representation



- Only $2 \log_2 N$ layers
  - By pairing terms
  - 2 layers per XOR

$$O = X_1 \oplus X_2 \oplus \cdots \oplus X_N$$

$$O = ((((X_1 \oplus X_2) \oplus (X_3 \oplus X_4)) \oplus ((X_5 \oplus X_6) \oplus (X_7 \oplus X_8))) \oplus (((\ldots$$

Learning the Network

- Next…