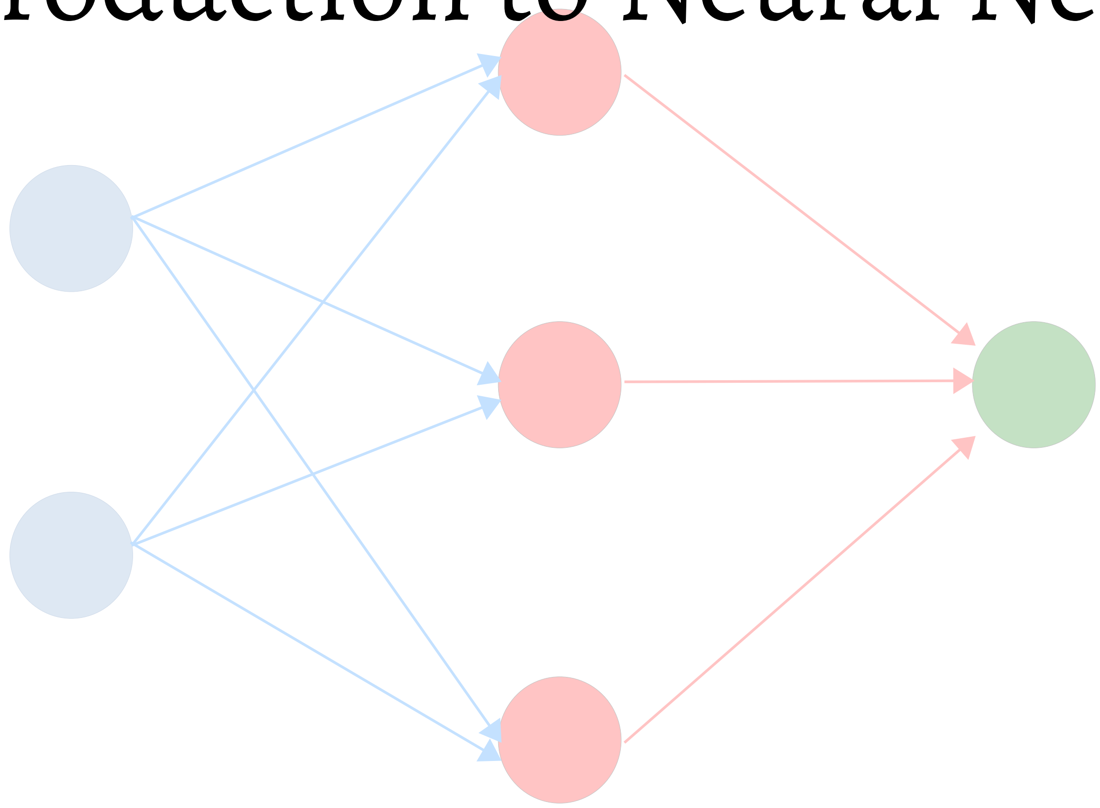
PH6232: Machine Learning for Physics applications

my Introduction to Neural Networks



Resources (content stolen shamelessly from...)
1. http://neuralnetworksanddeeplearning.com/
2. http://iamtrask.github.io/2015/07/12/basic-python-network/ and https://medium.com/technology-invention-and-more/how-to-build-a-simple-neural-network-in-9-lines-of-python-code-cc8f23647ca1
3. https://www.youtube.com/watch?v=bxe2T-V8XRs

Feb 3rd, 2023

Course webpage
https://alaha999.github.io/

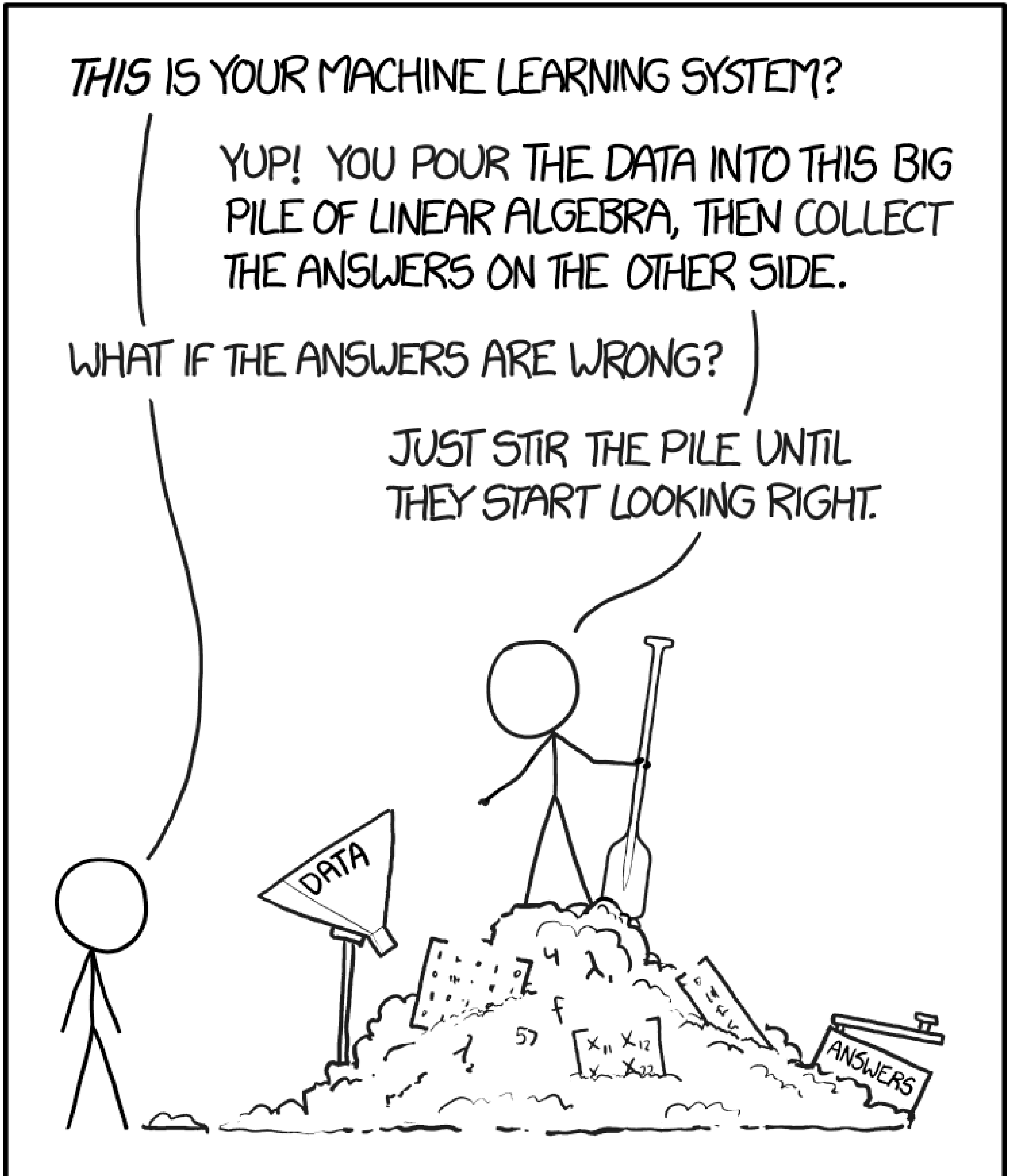
Aims

Working knowledge of neural networks ...
Working implies that we can implement code for a specific problem

Working knowledge of language of machine learning ...
Enabling us to read ideas in literature and critically analyse them

As we work through the course, discussions and problems will help build intuition and understanding of how to formulate a problem to use machine learning.

Assuming you are like Jon Snow ("know nothing")
Best way to learn is to do, especially NNs


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Setup

Basic Neural networks do classification, and regression

Classification: Is this a cat or dog?

Learn by looking at examples

Regression: How many marks will you get if you study for 5 hours, and play for 2 hours?

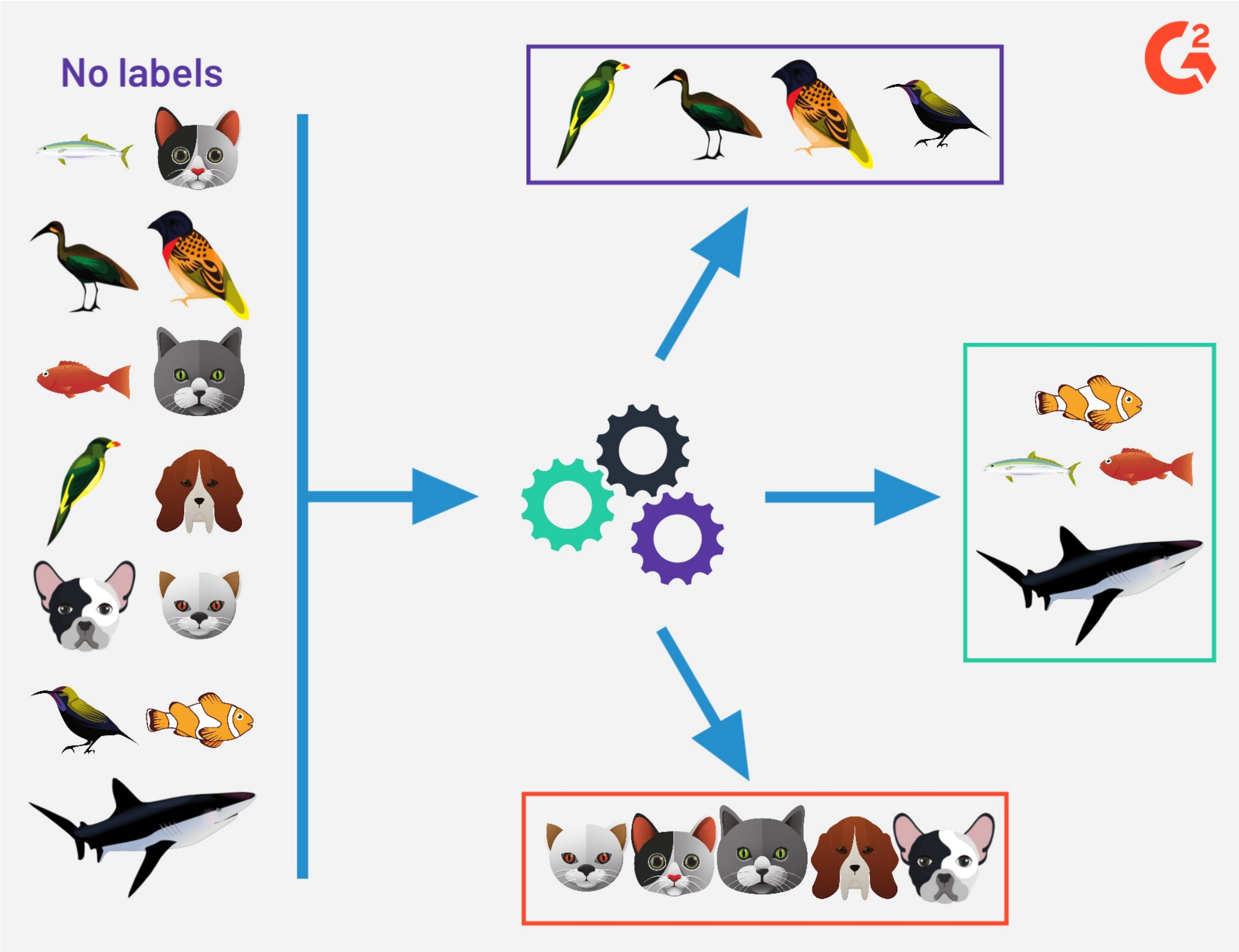
!!BSc final exam!!

Learn by looking at examples

10th Std.	FYBSc
11th Std.	SYBSc
12th Std.	

This is supervised learning....

This is unsupervised learning….



https://www.g2.com/articles/supervised-vs-unsupervised-learning

Iris flower data set

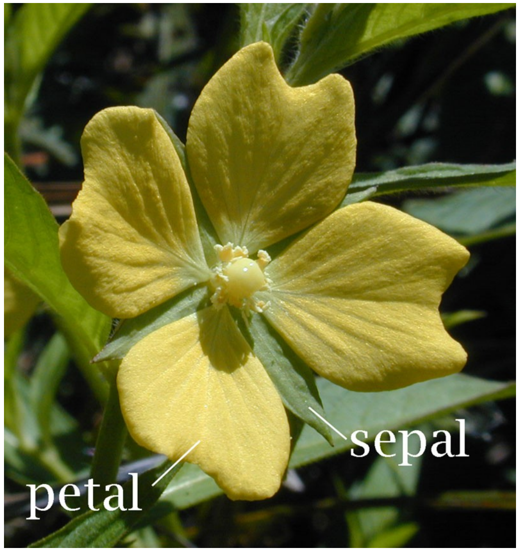
https://en.wikipedia.org/wiki/Iris_flower_data_set

Given sepal length, sepal width, petal length can we predict which flower it is?

Iris flower data set

https://en.wikipedia.org/wiki/Iris_flower_data_set

Given sepal length, sepal width, petal length can we predict which flower it is?

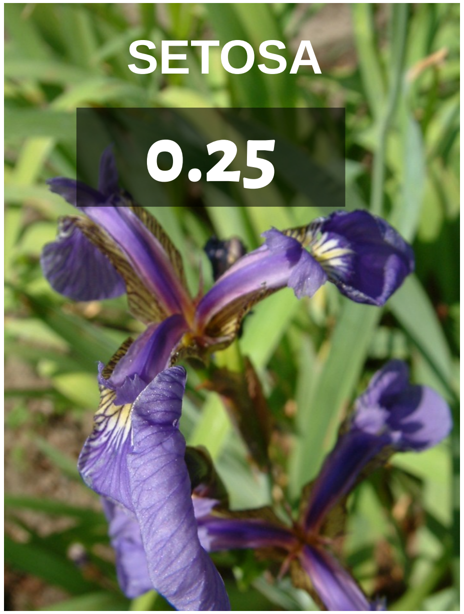


https://en.wikipedia.org/wiki/Sepal

Iris flower data set

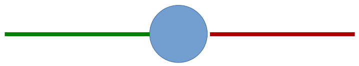
Given sepal length, sepal width, petal length can we predict which flower it is?

```
 SL , SW, PL,
5.1, 3.5, 1.4,
4.9, 3.0, 1.4,
6.2, 3.4, 5.4,
5.9, 3.0, 5.1
```



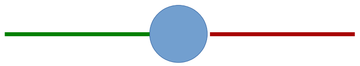
I assigned the numerical values to the labels at random....

Neurons

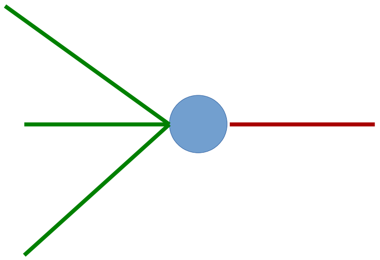


Given an input, there is an output

Neurons



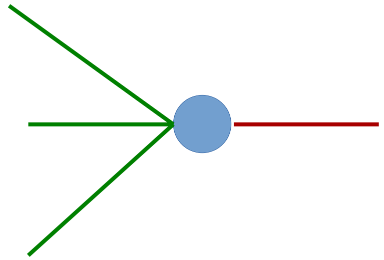
Given an input, there is an output $y = f(x)$



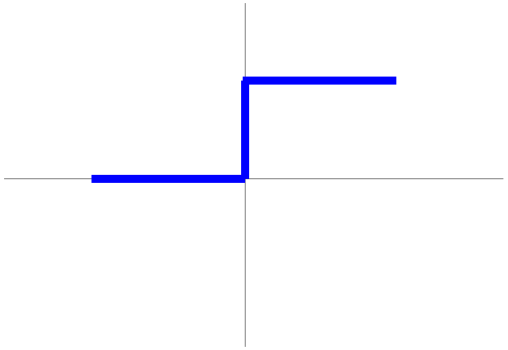
Given some inputs, there is an output $y = f(x_i)$

For eg. given how much sleep you had yesterday, how much you worked yesterday, and how much homework you have, you will decide if you want to come to a movie with me today.

Types of neuron



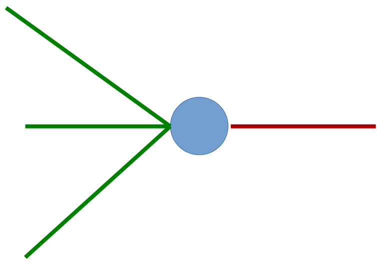
Perceptron



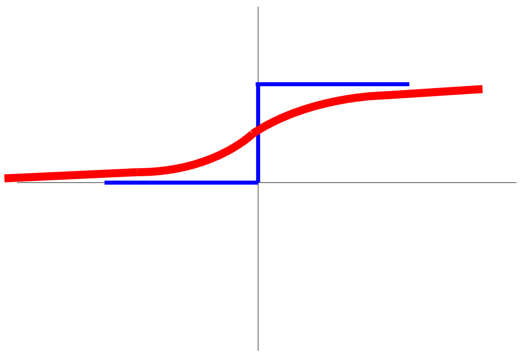
Binary output, either 0 or 1.

Thus, for small changes in input values, the neuron can suddenly fire (change state from 0 to 1).

Types of neuron



Sigmoid neuron



Smoother output

Thus, for small changes in input values, the neuron can have small changes in output.

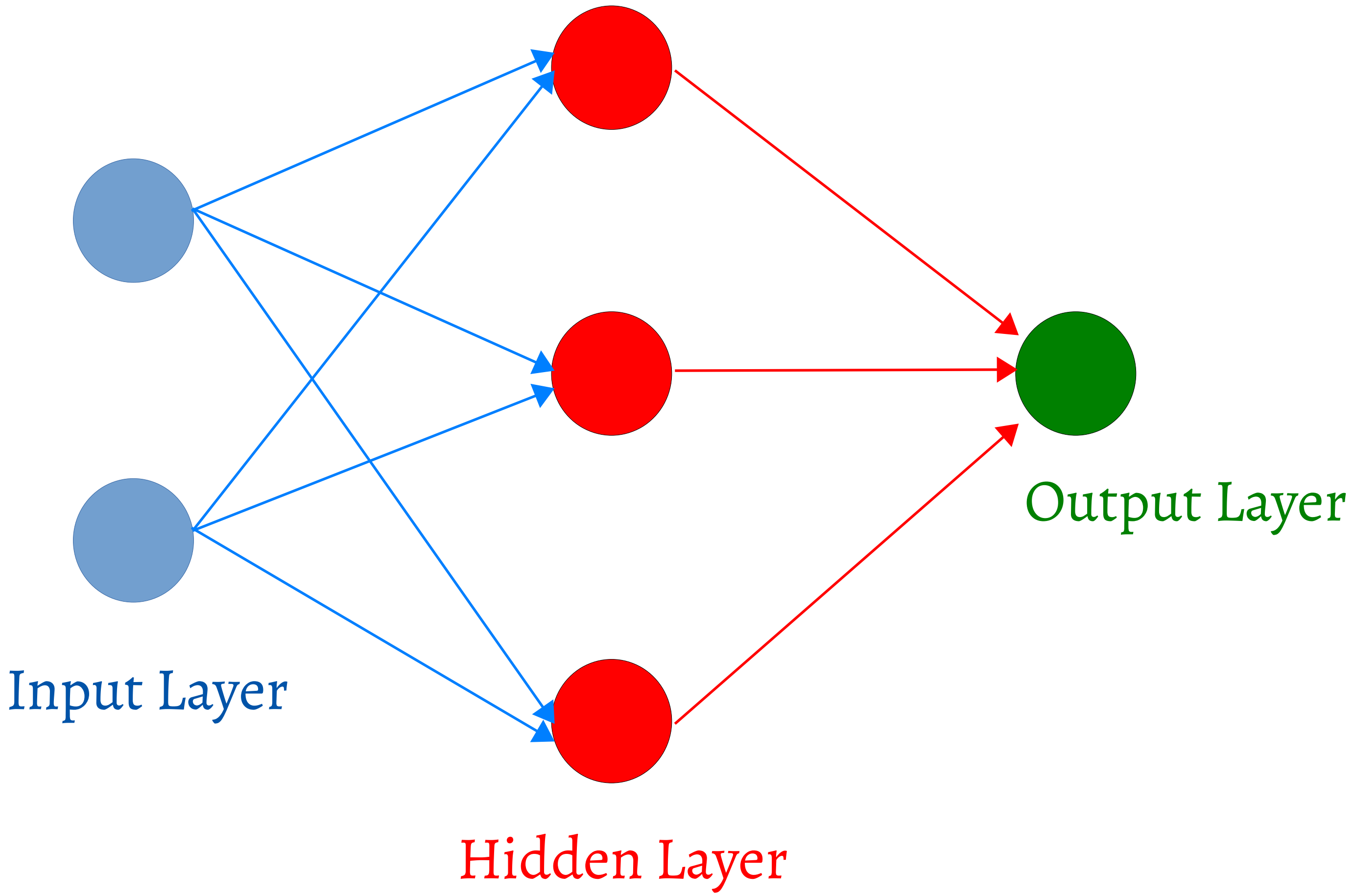
Types of neuron

Perceptron: output $y = w_i x_i + b$, turned to binary

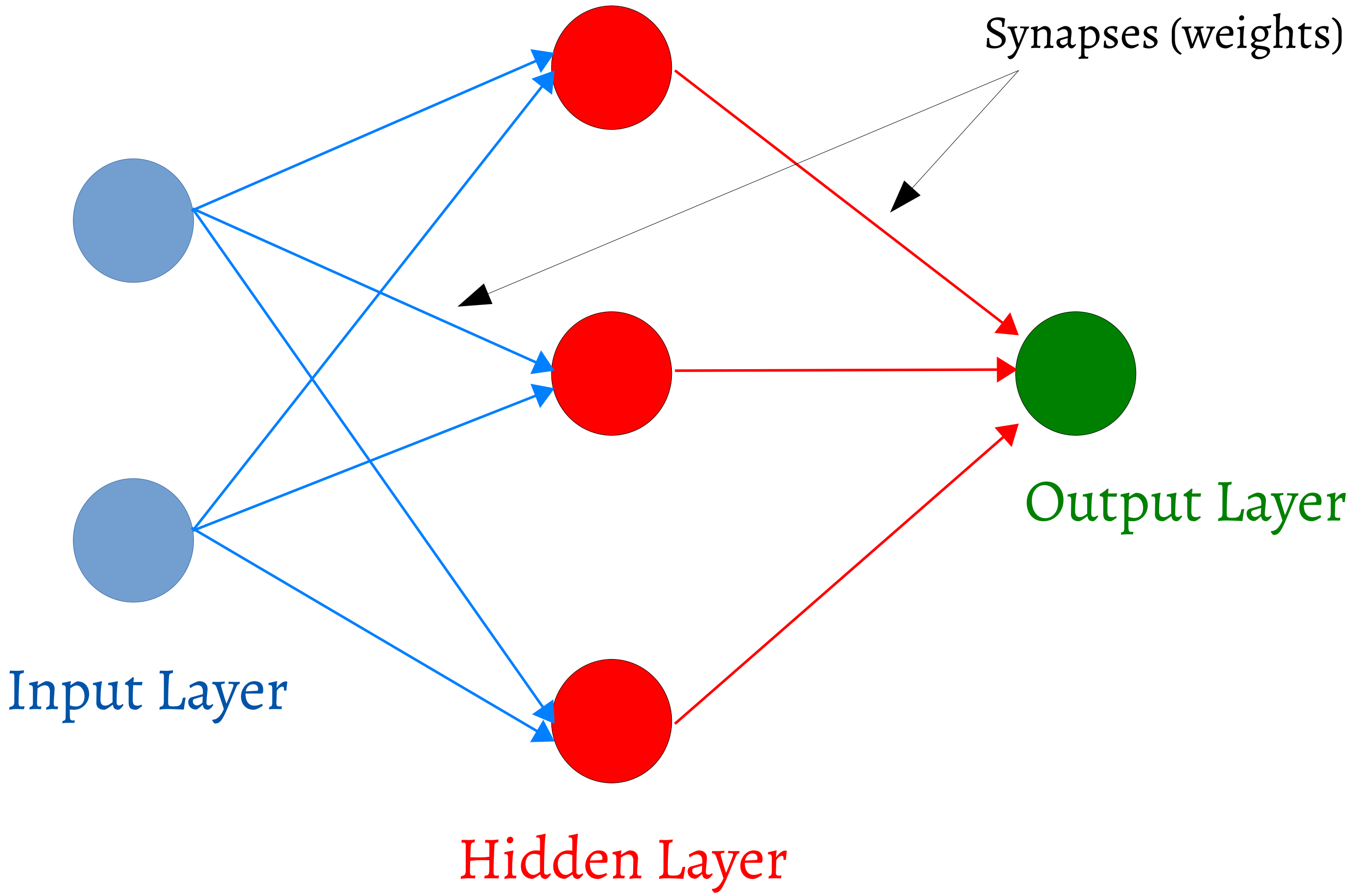
Sigmoid neuron: output $y = \sigma(z)$ where $z = x_i w_i + b$

And the sigmoid function : $\sigma(z) = \dfrac{1}{1 + e^{-z}}$

An example network: nomenclature

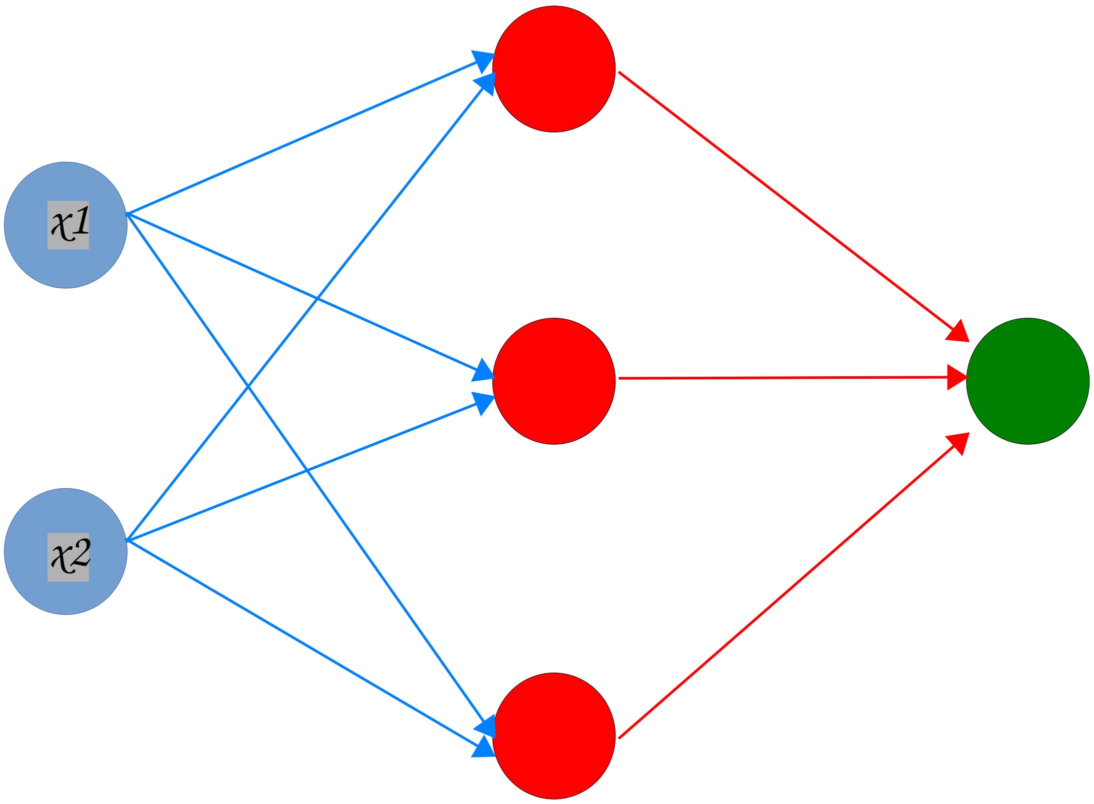


An example network: nomenclature



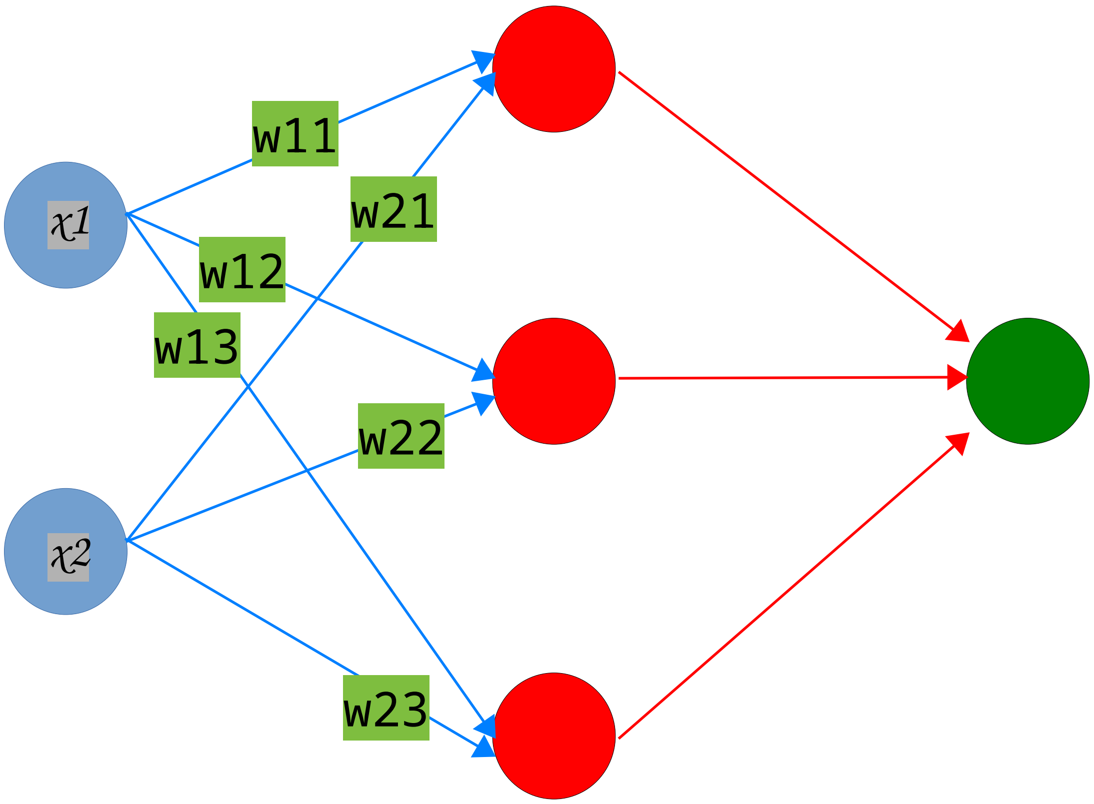
Can have more than one hidden layer… deep learning, deep networks.

An example network



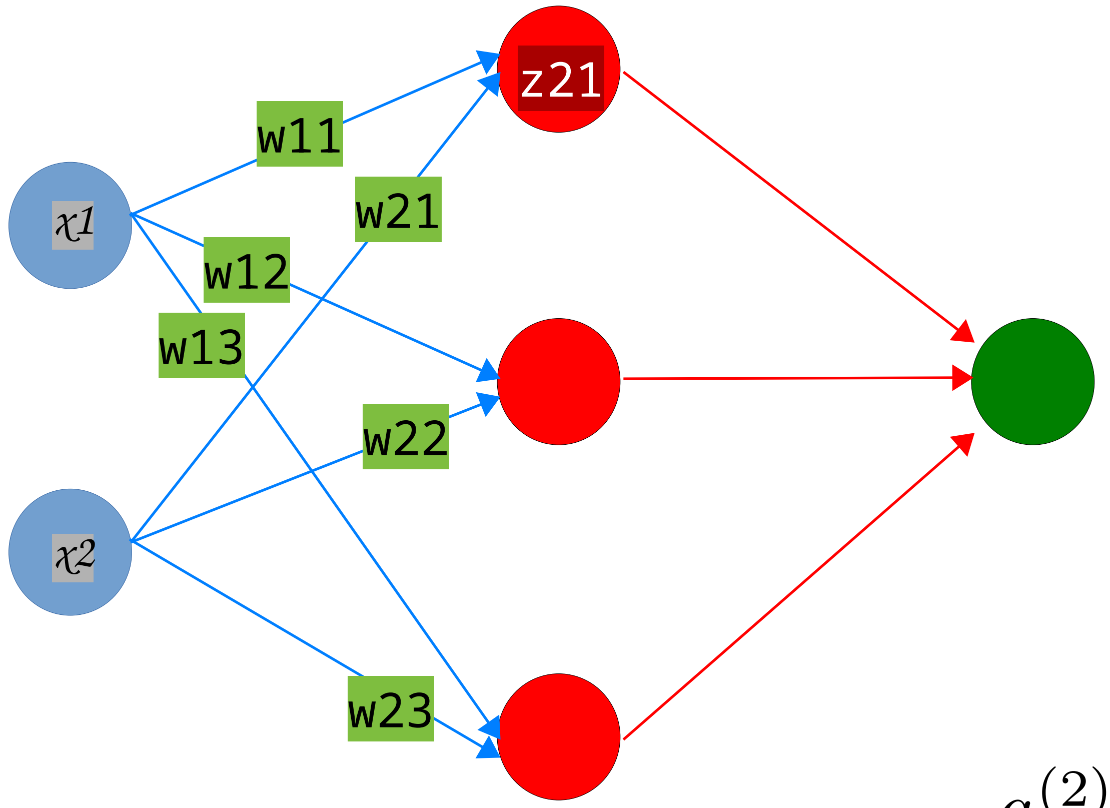
There is some known input data X, for which known output is Y (we want to train our network such that given some X in future, the output can be predicted)

An example network



An example network

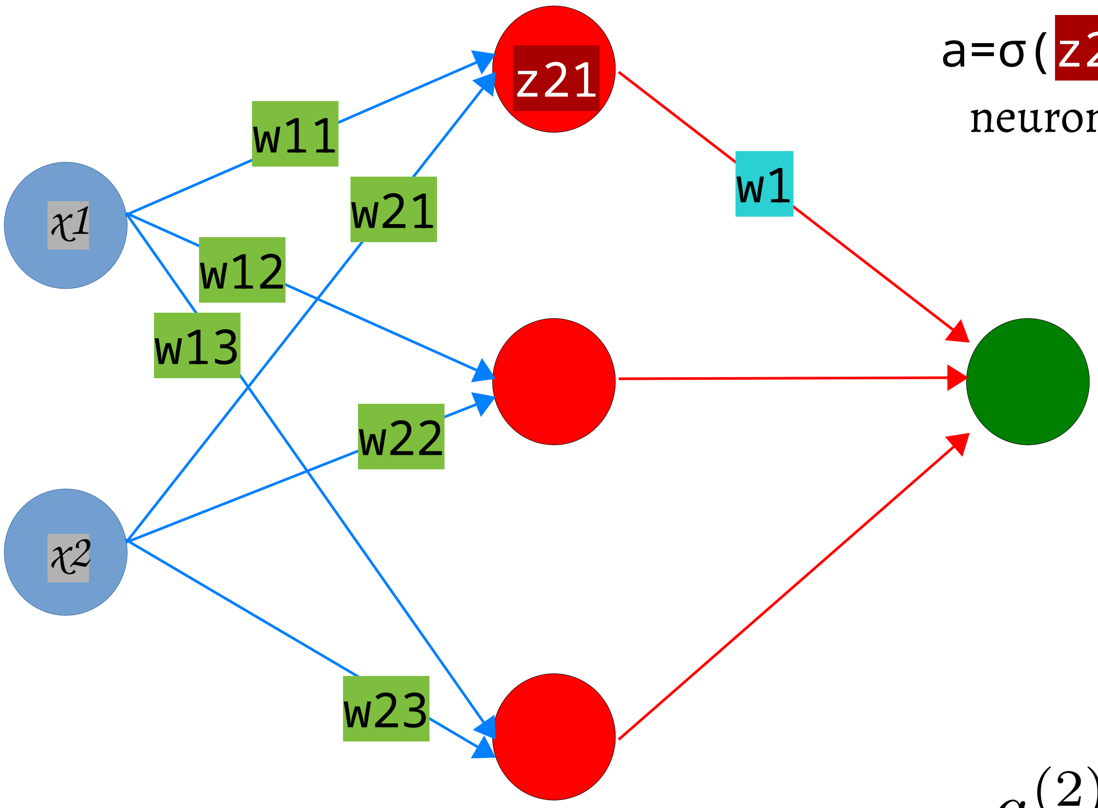
z21 = w11 $x1$ + w21 $x2$



$$z^{(2)} = XW^{(1)}$$
$$a^{(2)} = \sigma(z^{(2)})$$

An example network

z21 = w11 $x1$+w21 $x2$

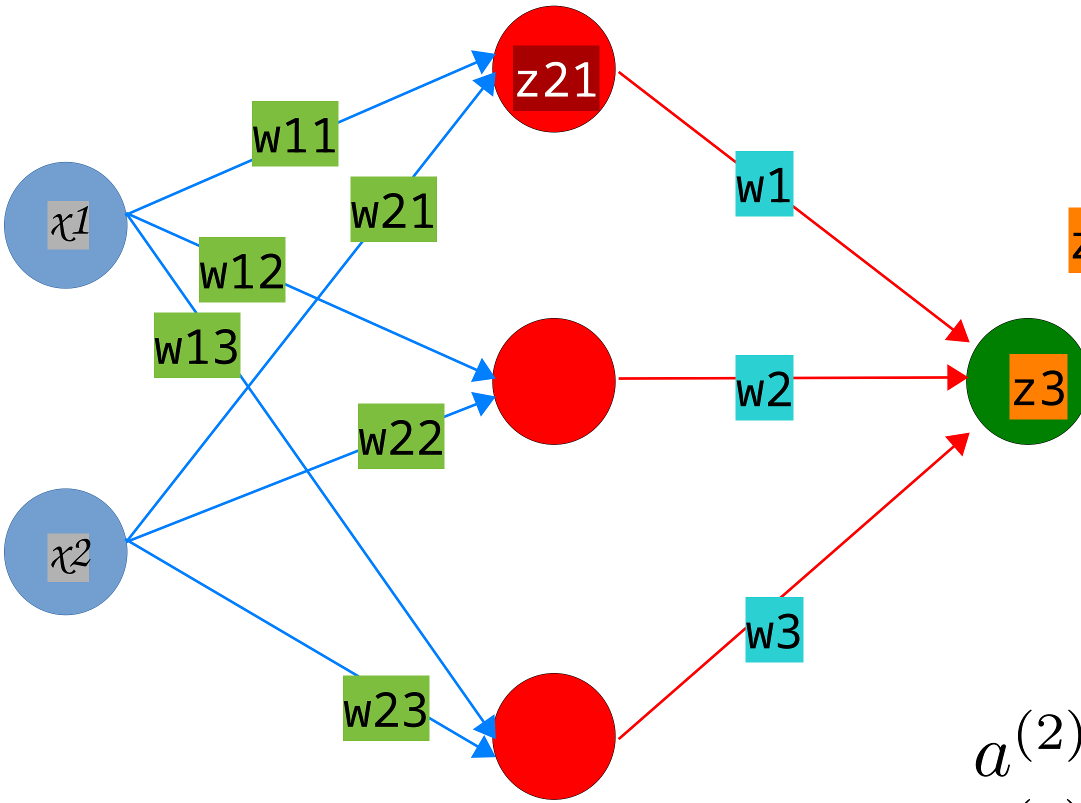


a=σ(z21) is activation of neuron 1 in hidden layer

$$z^{(2)} = XW^{(1)}$$
$$a^{(2)} = \sigma(z^{(2)})$$

An example network

z21 = w11 $x1$ + w21 $x2$,



z3 = σ(z21) w1 +
σ(z22) w2 +
σ(z23) w3

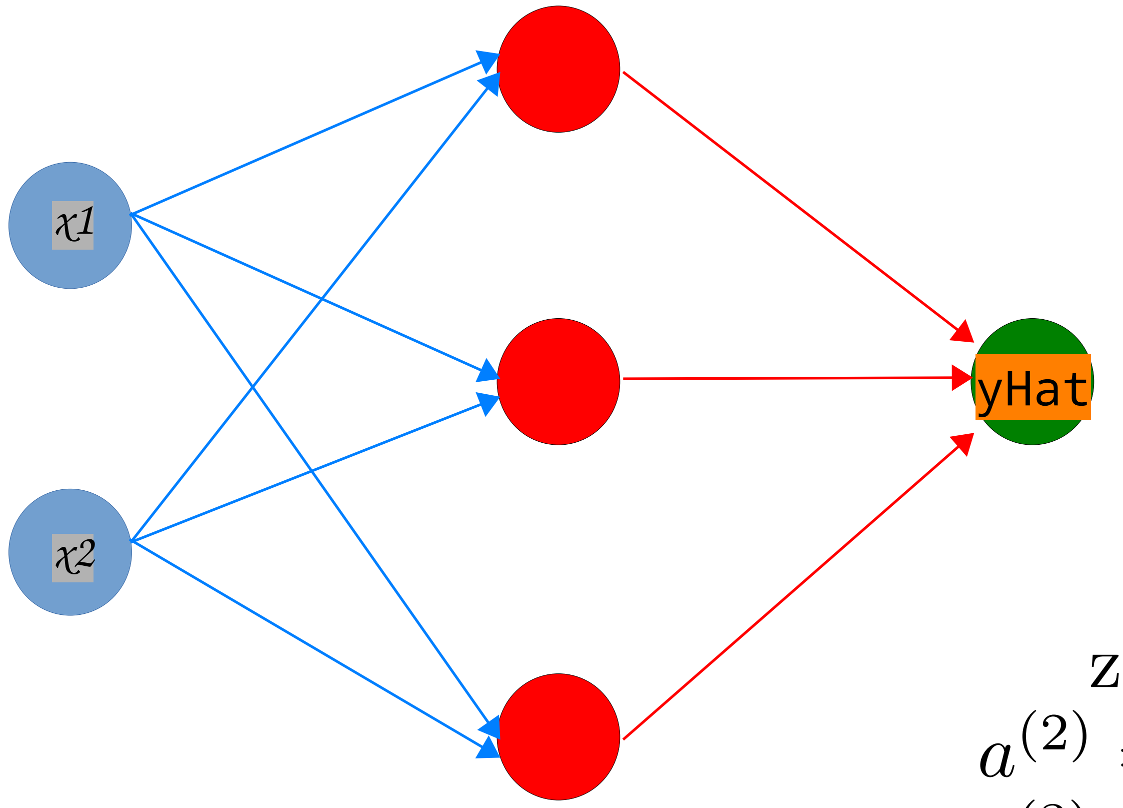
$$z^{(2)} = XW^{(1)}$$
$$a^{(2)} = \sigma(z^{(2)})$$
$$z^{(3)} = a^{(2)}W^{(2)}$$

Output of network

$$\hat{y} = \sigma(z^{(3)})$$

Forward propagation

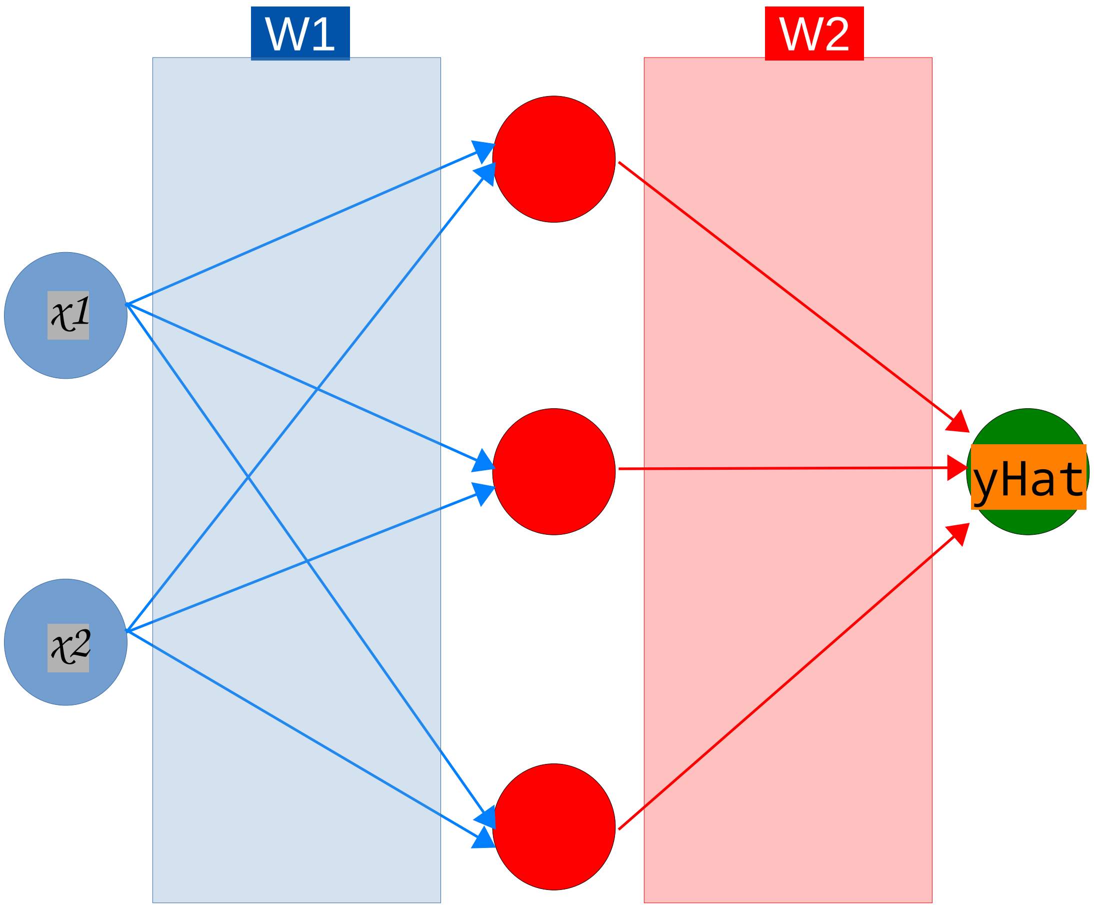
Given inputs, and some weights, what is the output.



$$z^{(2)} = XW^{(1)}$$
$$a^{(2)} = \sigma(z^{(2)})$$
$$z^{(3)} = a^{(2)}W^{(2)}$$

Output of network → $\hat{y} = \sigma(z^{(3)})$

Gradient descent



What combination of weights gives the best predictions?

Cost function

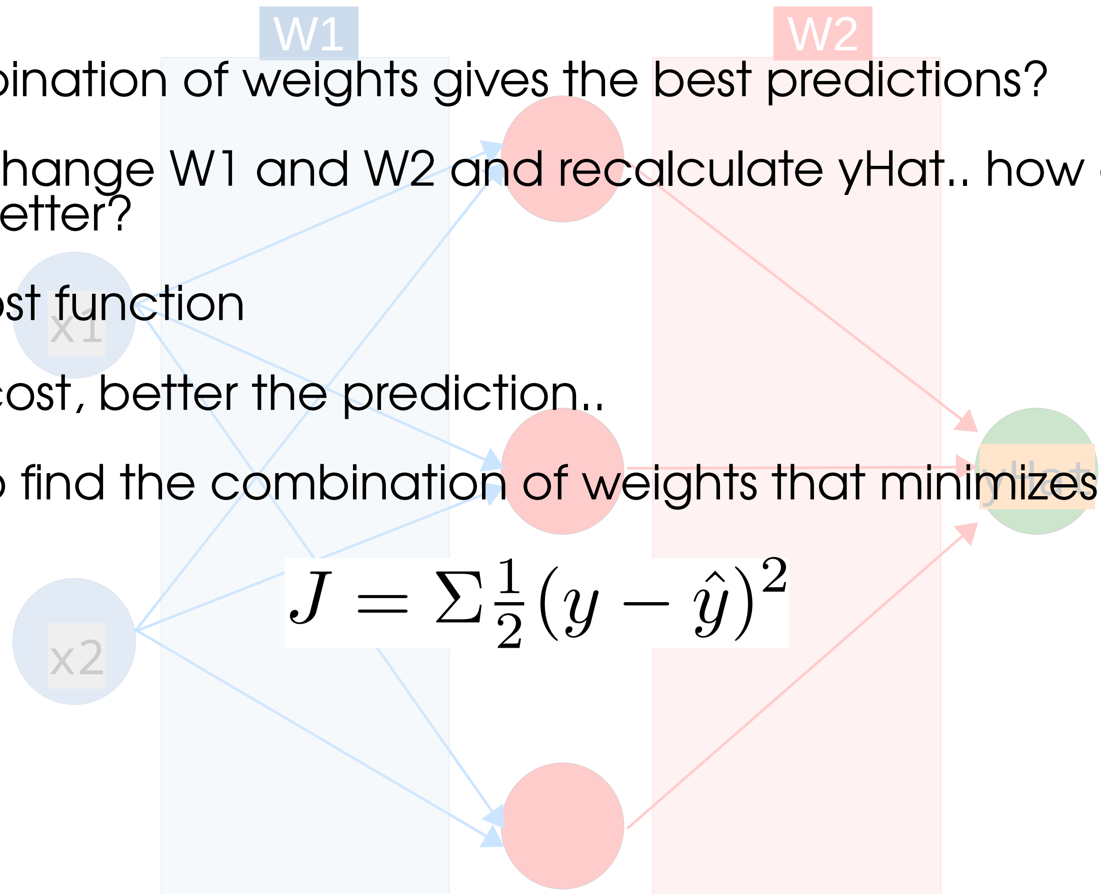
What combination of weights gives the best predictions?

We could change W1 and W2 and recalculate yHat.. how do we know if its better?

Define a cost function

Lower the cost, better the prediction..

We need to find the combination of weights that minimizes the cost.



$$J = \Sigma \frac{1}{2}(y - \hat{y})^2$$

Gradient descent

We have cost as a function of the weights.

We calculate the gradient of cost w.r.t. weights, identifying the direction in which J is reducing.

$$\frac{\partial J}{\partial W_1}, \frac{\partial J}{\partial W_2}$$

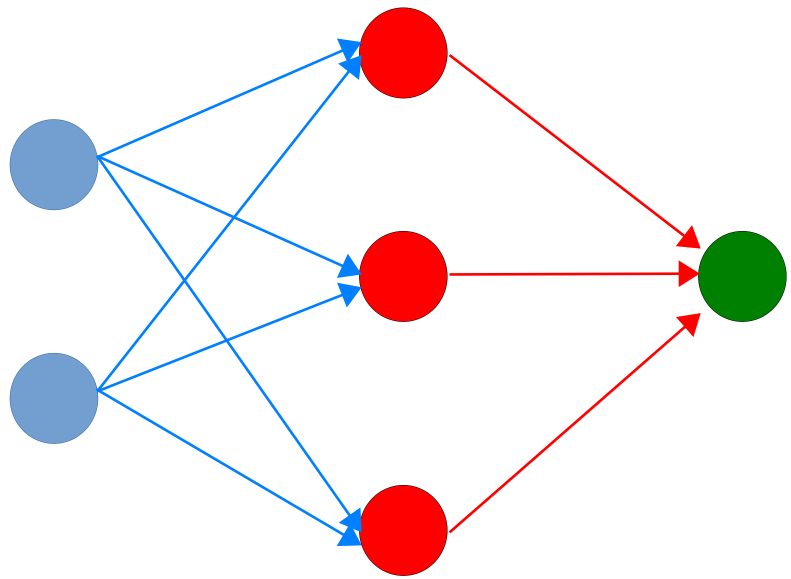
We change the weights by some amount (Rate $\times$ Gradient), and recalculate cost.

And thus iterate until we are satisfied with the weights...

And that's it.

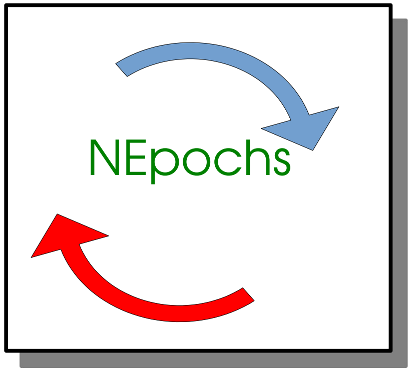
Forward propagate with weights

Change the weights using the gradient of cost.



Calculate the final estimate and calculate the cost.

Back propagate using gradients



Math of back propagation

1

$$
\begin{aligned}
Z_2 &= X \cdot W_1 \\
a_2 &= \sigma(Z_2) \\
Z_3 &= a_2 \cdot W_2 \\
\hat{y} &= \sigma(Z_3) \\
&= \sigma(a_2 \cdot W_2) \\
&= \sigma[\sigma(X \cdot W_1) \cdot W_2]
\end{aligned}
$$

2

$$
\begin{aligned}
J &= \Sigma \tfrac{1}{2}(y - \hat{y})^2 \\
\frac{\partial J}{\partial W_2} &= -(y - \hat{y}) \ \frac{\partial \hat{y}}{\partial W_2} \\
\frac{\partial J}{\partial W_1} &= -(y - \hat{y}) \ \frac{\partial \hat{y}}{\partial W_1}
\end{aligned}
$$

3

$$
\frac{\partial \hat{y}}{\partial W_2} = \sigma'(a_2 \cdot W_2)\, a_2 \quad \text{using}(5)
$$

$$
\boxed{
\begin{aligned}
\frac{\partial J}{\partial W_2} &= a_2^T \cdot \delta^3 \\
\text{where} \quad \delta^3 &= -(y - \hat{y}) \ \sigma'(a_2 \cdot W_2) = -(y - \hat{y}) \ \sigma'(Z_3)
\end{aligned}
}
$$

4

$$
\begin{aligned}
\frac{\partial \hat{y}}{\partial W_1} &= \sigma'(Z_3) \frac{\partial Z_3}{\partial W_1} \\
&= \sigma'(Z_3) \frac{\partial a_2}{\partial W_1} W_2 \\
&= \sigma'(Z_3) W_2^T \sigma'(Z_2) \cdot X
\end{aligned}
$$

$$
\boxed{\frac{\partial J}{\partial W_1} = X^T \ \delta^3 \ W_2^T \ \sigma'(Z_2)}
$$

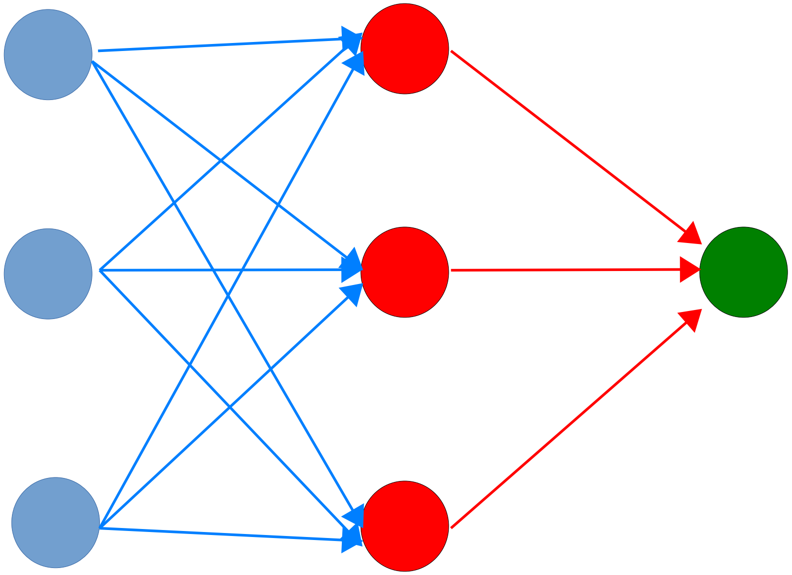
My implemented network

Iris data

3 variables

3 hidden layer neurons

4 training events



Dimensions
$X = 4 \times 3$
$W_1 = 3 \times 3$
$W_2 = 3 \times 1$

$Z_2, A_2 = 4 \times 3$

$yHat = 4 \times 1$

$dJ/dW_1 = 3 \times 3$
$dJ/dW_2 = 3 \times 1$