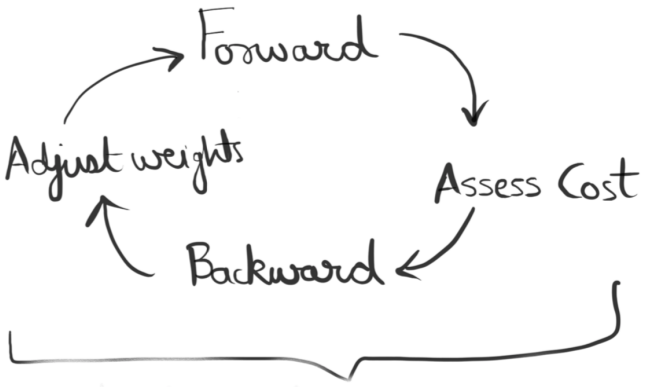
5 Components of Neural Network Training

1. Training Loop
2. Forward
3. Cost
4. Backward
5. Adjust Weights



Just plain calculations: +, - , x, / , functions

No Algorithmic Cleverness

No smart data structures

1. Training Loop

Epochs and Batch Size

Problem 1: We are training on a dataset of 506 houses, with batch size of 10 and for 20 epochs, How many times weight update will happen?

2. Forward

Data

x1	x2		y
1	2		0
1	3		0
2	1		0
2	3		0
3	1		0
1	4		1
4	2		1

fig 2a

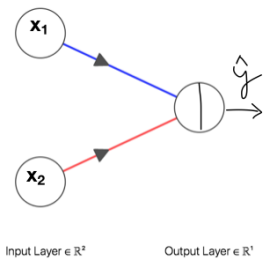
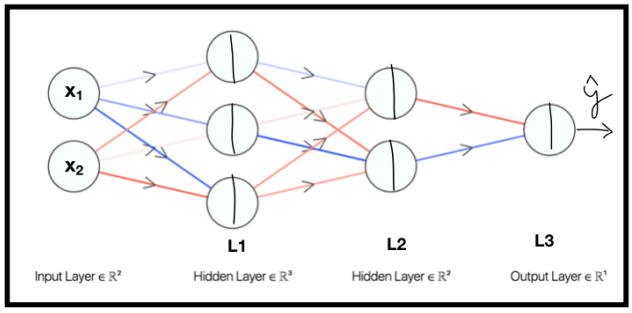


fig 2b



Problem 1: For the given single neuron setup in fig 2a, predict the output for the first input batch of size 4.

Problem 2: For the given 2 matrices A and B,

A

$$\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 3 & -3 \end{bmatrix}$$

B

$$\begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 3 & 1 & 3 \end{bmatrix}$$

- Compute A X B (Matrix multiplication)
- Compute elementwise Sigmoid on the output of above multiplication

Problem 3: For the given neural network in fig 2b,

- How many weights are there in first hidden layer?
- How many weights are there in total?
- Compute the activation from the first hidden layer of the network for the first input batch of size 4?
- How many activations are produced from first hidden layer for single input sample?

3. Cost

- Mean Squared Error for regression problems

$$J = \frac{1}{2N}\sum_{i=0}^{N}(y - \hat{y})^2$$

- Categorical Cross Entropy Error for classification problems

$$J = -\frac{1}{N}\sum_{i=0}^{N}\left(y^{(i)}\log\left(\hat{y}^{(i)}\right) + (1 - y^{(i)})\log\left(1 - \hat{y}^{(i)}\right)\right) \tag{13}$$

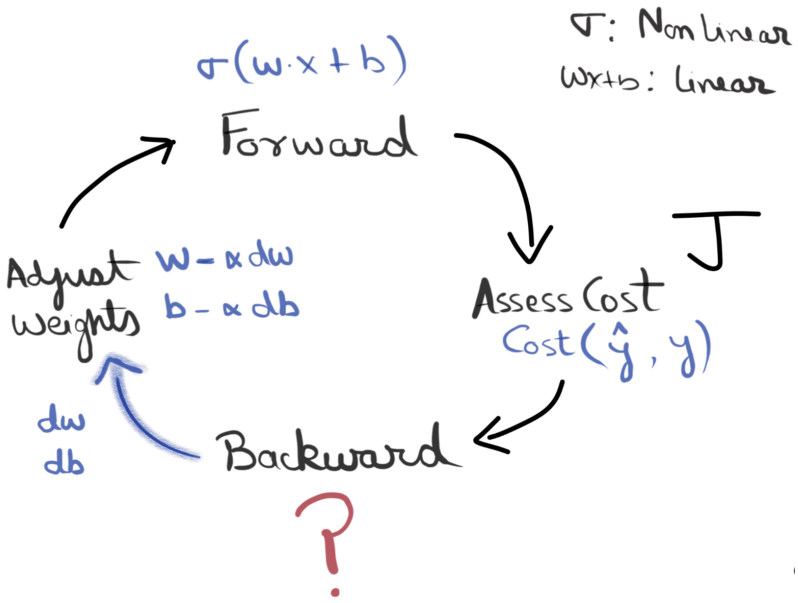
Problem 1 : For the single neuron setup like above compute the Mean Squared Error (MSE) cost for the first batch of size 3 given the following training data. And Weights as w1 = -0.2, w2 = 0.5

x1	x2	y
20	16	5
40	20	3
10	6	4
20	12	4

Problem 2 : For the following values of Y and Y_hat for a classification problem, Compute Categorical Cross entropy error

y	y_hat
1	0.96
0	0.003
0	0.8
1	0.88

4. Backward



Derivative from first principles

Problem 1: Compute derivative of following functions from first principles

$y = 5x$

$y = x^2$

Simple Derivative Problems

Problem 1: Find derivatives for following functions

$y = \sin(x)$

$y = x + 2$

$y = 5\sin(x)$

$y = 3x$

$y = 5\sin(x) + x^4$

$y = 5x^4\sin(x)$

$y = 2e^x\log(x)$

Derivatives Chain Rule

Problem 1: Find derivatives for following functions using chain rule

$$y = \log(x^2)$$

$$y = \sin^2(x)$$

$$y = \sin(x^2)$$

$$y = e^{5x^2}$$

$$y = \frac{1}{1 + e^{-10x+3}}$$

Gradient Descent

Problem 1: Tale of 2 functions: First is *f*, the function we are trying to predict. second is *J*, the cost function. Whats your general observation about *f* and *J* ?

Problem 2: Simply stated gradient descent means that move in the direction of negative slope in some proportion to the incline of the slope.

```
- Which of above 2 functions we would do gradient descent?
- What is a learning rate?
```

Problem 3: Given the following cost function J

$$J(x) = (x - 3)^2$$

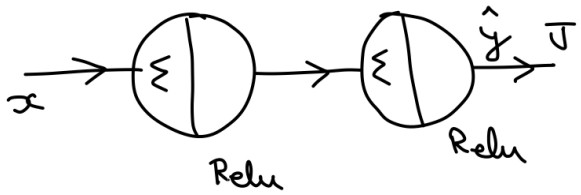
and our current weight is 7, given the learning rate of 0.1

- What will be the length of our next stride?
- What will be the new weight?

Backpropagation

Problem 1: What is being back propagated?

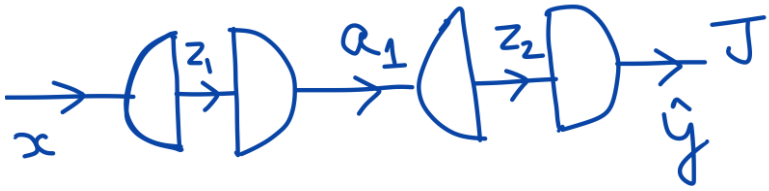
Problem 2: In the following setup




```
 - How many functions are involved?
 - How many weights are involved?
 - With only one training sample x=3, y=2 and weights w1=0.5, w2=0.8 Whats y
_hat?
 - Whats the MSE Cost?
```

Problem 3:

Just expanded the above neural network to show weighted sum and activation components



To go backward, compute the following for previously given weights and training sample

$$\frac{dJ}{d\hat{y}} \quad \frac{d\hat{y}}{dz2} \quad \frac{dz2}{dw2} \quad \frac{dz2}{da1} \quad \frac{da1}{dz1} \quad \frac{dz1}{dw1}$$

Problem 4:

Whole purpose of back propagation or this notebook is to compute ***how cost changes with weights*** i.e

$$\frac{dJ}{dweights}$$

What are the values of:

a) $\frac{dJ}{dw1}$

b) $\frac{dJ}{dw2}$

(Hint: Use Chain Rule & Problem 3 results)

5. Update Weights

Problem 1: In the above problem after you have computed $\frac{dJ}{dw1}$ and $\frac{dJ}{dw2}$

what are new weights if the learning rate is 0.1

Forward Again

Problem 1 : Plug in those new weights and check if the cost is reduced.