CS 307

NEURAL NETWORKS

WHAT <u>IS</u> A "NEURAL NETWORK"?

- ARTIFICAL "ANN" / "NN"
- A FUNCTION
  - OF INPUT DATA AND PARAMETERS
  - OFTEN REPRESENTED AS A NETWORK
- PARAMETERS LEARNED FROM DATA

↑

"WEIGHTS"

Logistic Regression

Sigmoid Function $\sigma(x) = \dfrac{1}{1+e^{-x}}$ $\qquad \sigma : \mathbb{R} \to [0,1]$

$$p(x) = P[Y=1 \mid X = x]$$

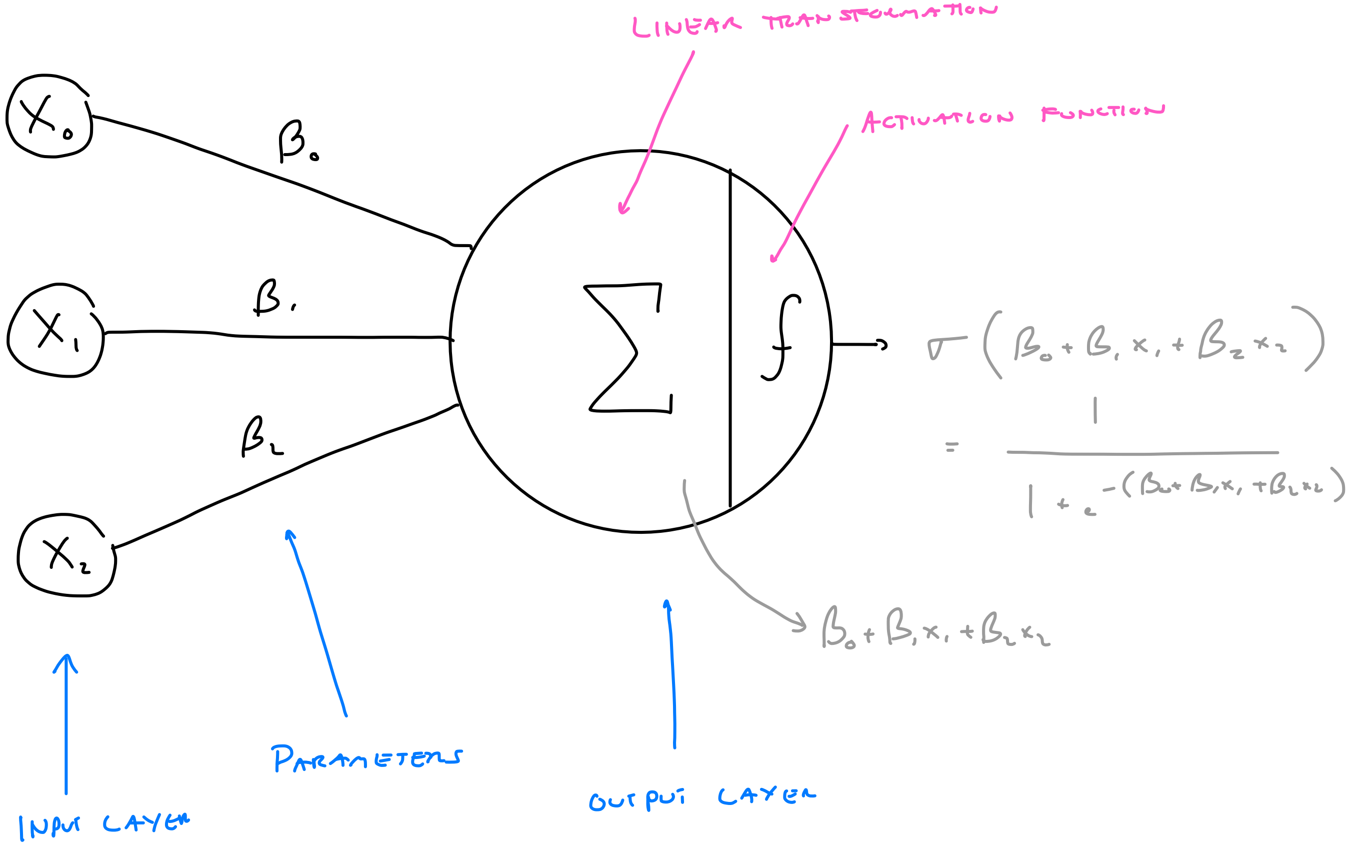
THREE PARAMETERS

TWO INPUT FEATURES

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2}_{\text{"LOGIT"}}$$

LR AS A "NETWORK"

LET $x_0 = 1$

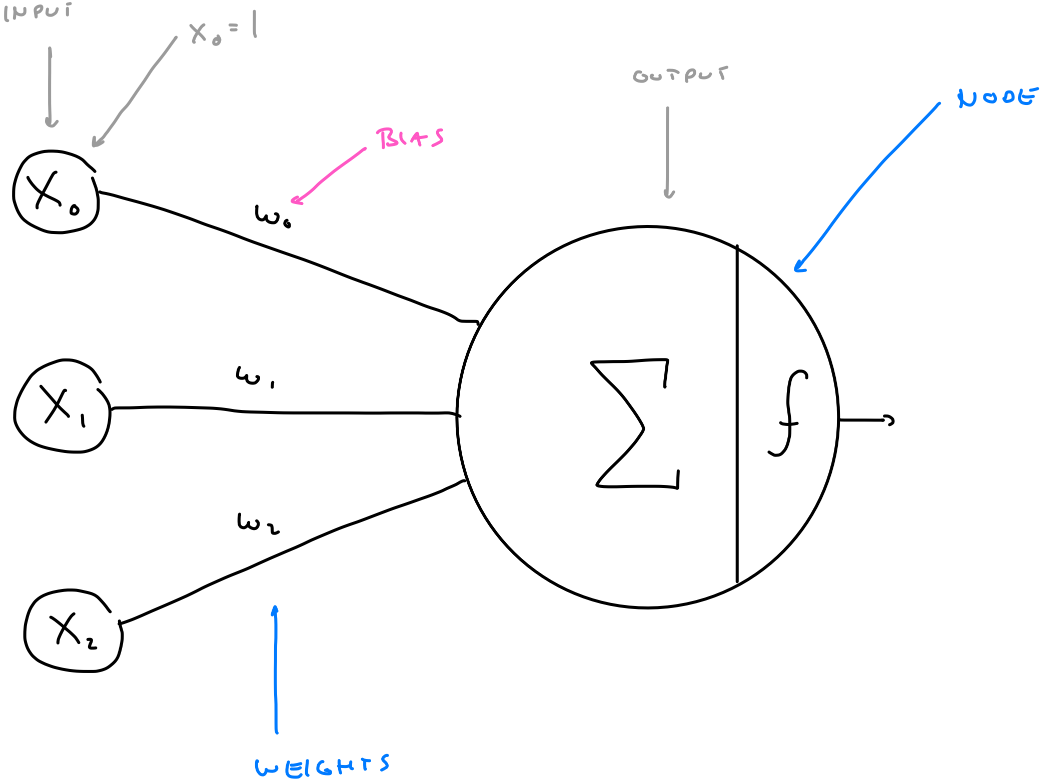


$$\sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2$$

NN NOMENCLATURE

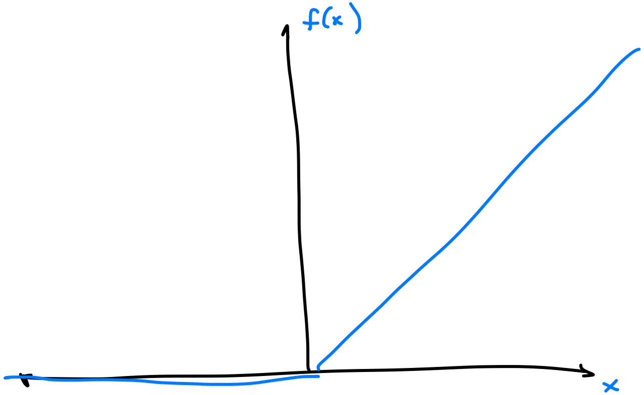


ReLU

Rectified Linear Unit

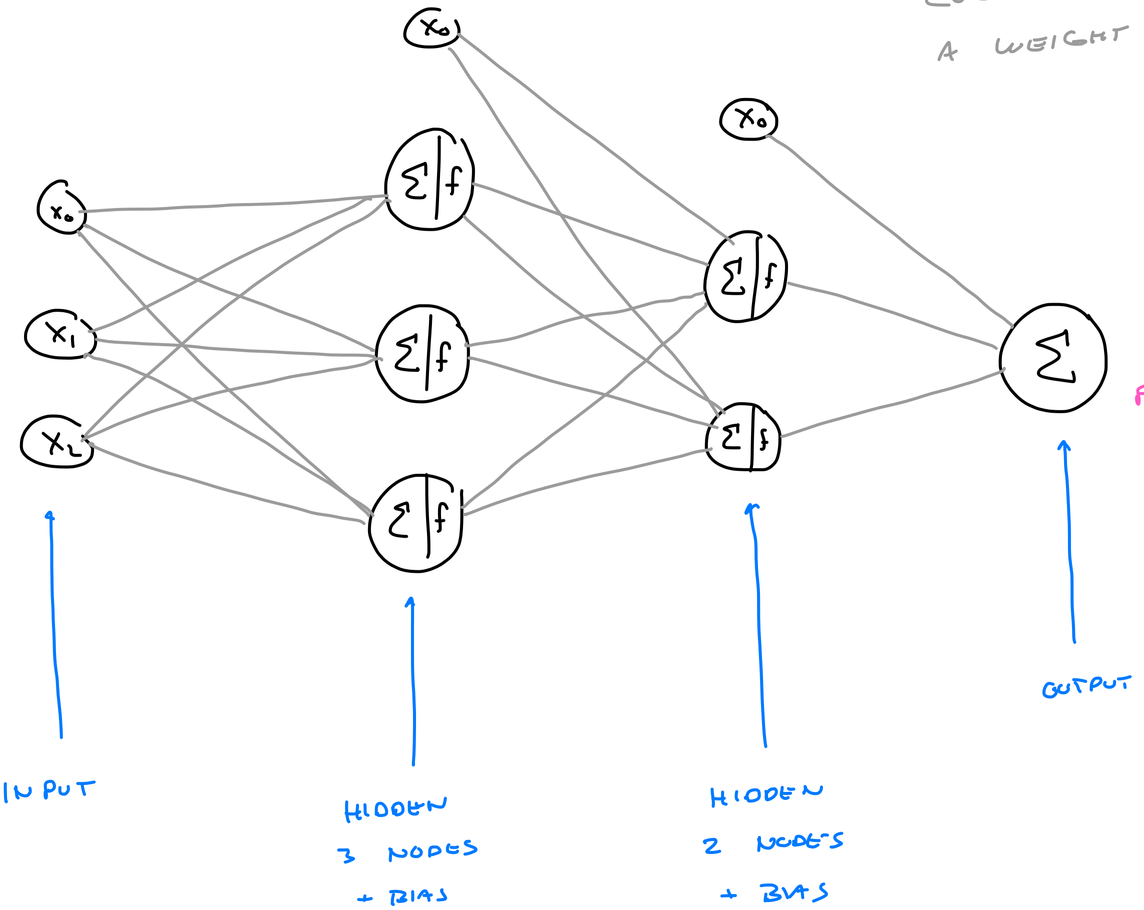
Activation

$$f(x) = \max(0, x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$



GOING 'DEEP"

EVERY CONNECTION REPRESENTS A WEIGHT PARAMETER



- SINGLE OUTPUT FOR BINARY
- ONE-HOT FOR MULTICLASS

WHAT DO WE CONTROL?

- HOW MANY LAYERS?
- HOW MANY NODES PER LAYER?

NETWORK ARCHITECTURE

- WHICH ACTIVATION FUNCTION?

USE RELU FOR HIDDEN

- HOW TO FIT / OPTIMIZE?
    - METHOD
    - LEARING RATE

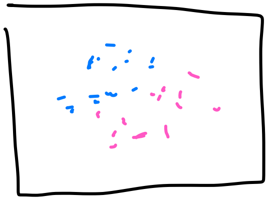
DEEP LEARNING = BIG NETWORKS

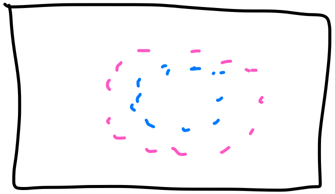
How to train NNs?

"How to learn weights from data?"

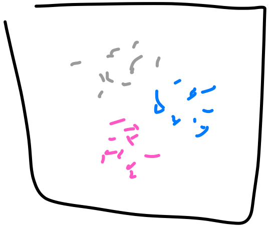
- Optimization problem!
  - Define loss function
  - Backpropagation (chain rule)
  - SGD / ADAM

1. LEARN TOY BINARY DATA





2. LEARN MULTICLASS DATA



3. LEARN TO CLASSIFY IMAGES!

Images as Feature Data

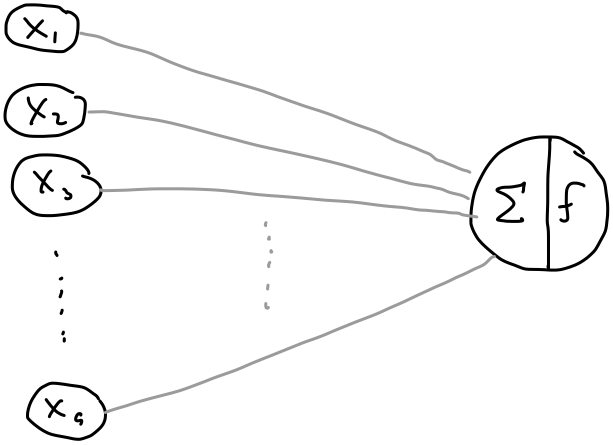
1	2	3
4	5	6
7	8	9

3x3 pixel image

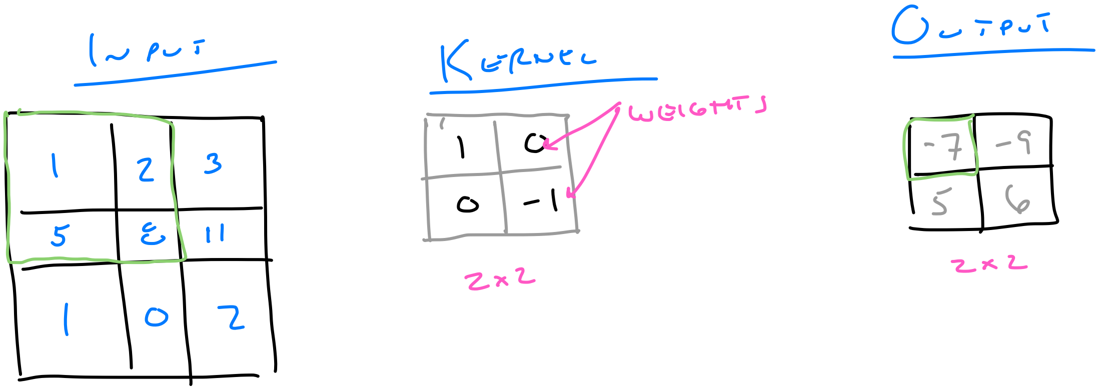
"Flatten"

x_1	x_2	x_3	....	x_9
1	2	3	....	9

For simplicity, assume greyscale images



Conv Net



1	2	3
5	8	11
1	0	2

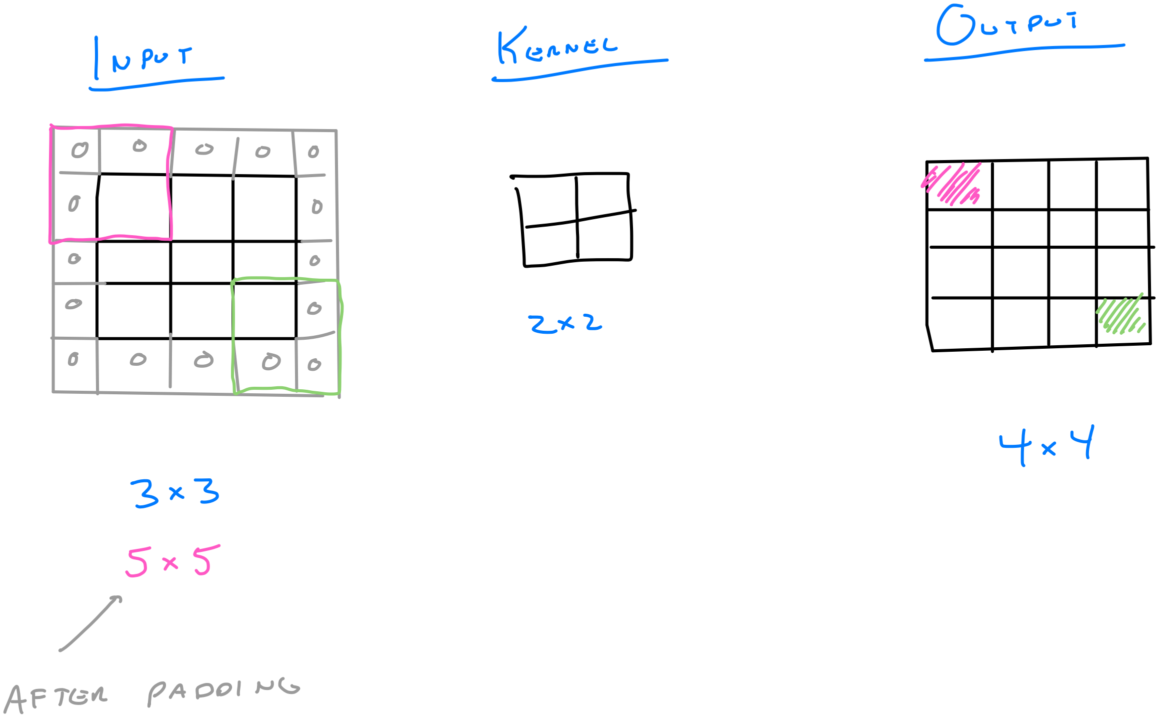
1	0
0	-1

-7	-9
5	6

3x3

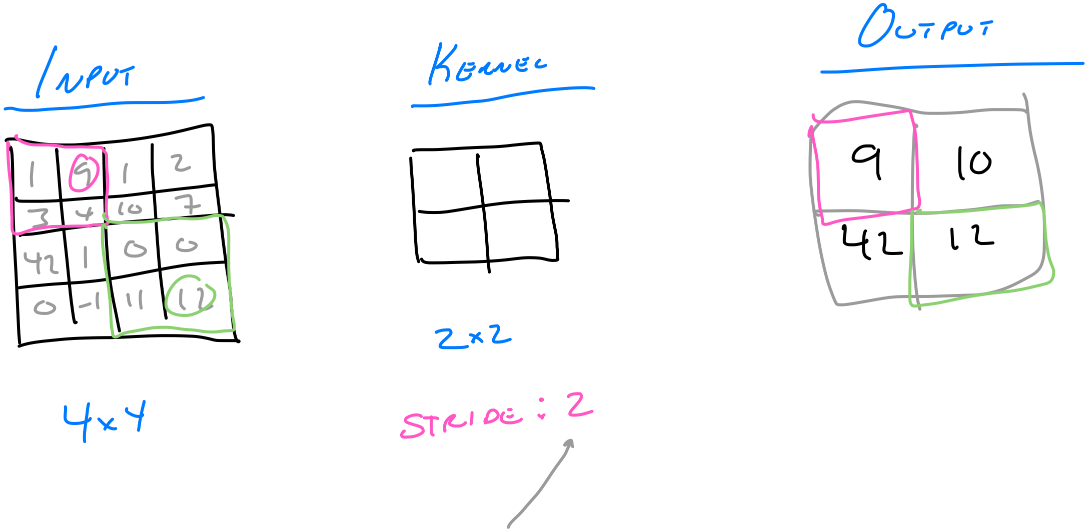
$$1(1) + 0(2) + 0(5) - 1(8) = -7$$

PADDING


INPUT
KERNEL
OUTPUT
0 0 0 0 0
0 0
0 0
0 0
0 0 0 0 0
2×2
4×4
3×3
5×5
AFTER PADDING

Pooling

[✓ MAX] / AVERAGE



Input

1	9	1	2
3	4	10	7
42	1	0	0
0	-1	11	12

4x4

Kernel

2x2

STRIDE: 2

Often set to kernel size

Output

9	10
42	12