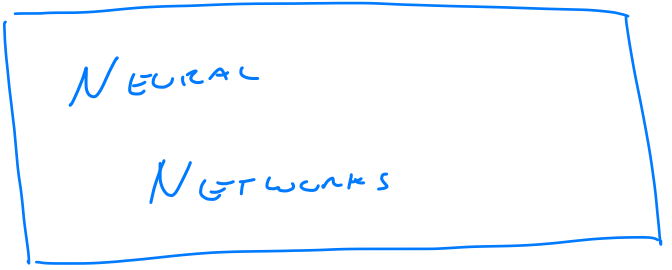
# CS 307

Spring 2024

Dalpiaz



# WHAT IS 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 $\sigma(x) = \frac{1}{1+e^{-x}}$ $\sigma : \mathbb{R} \to [0,1]$

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

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

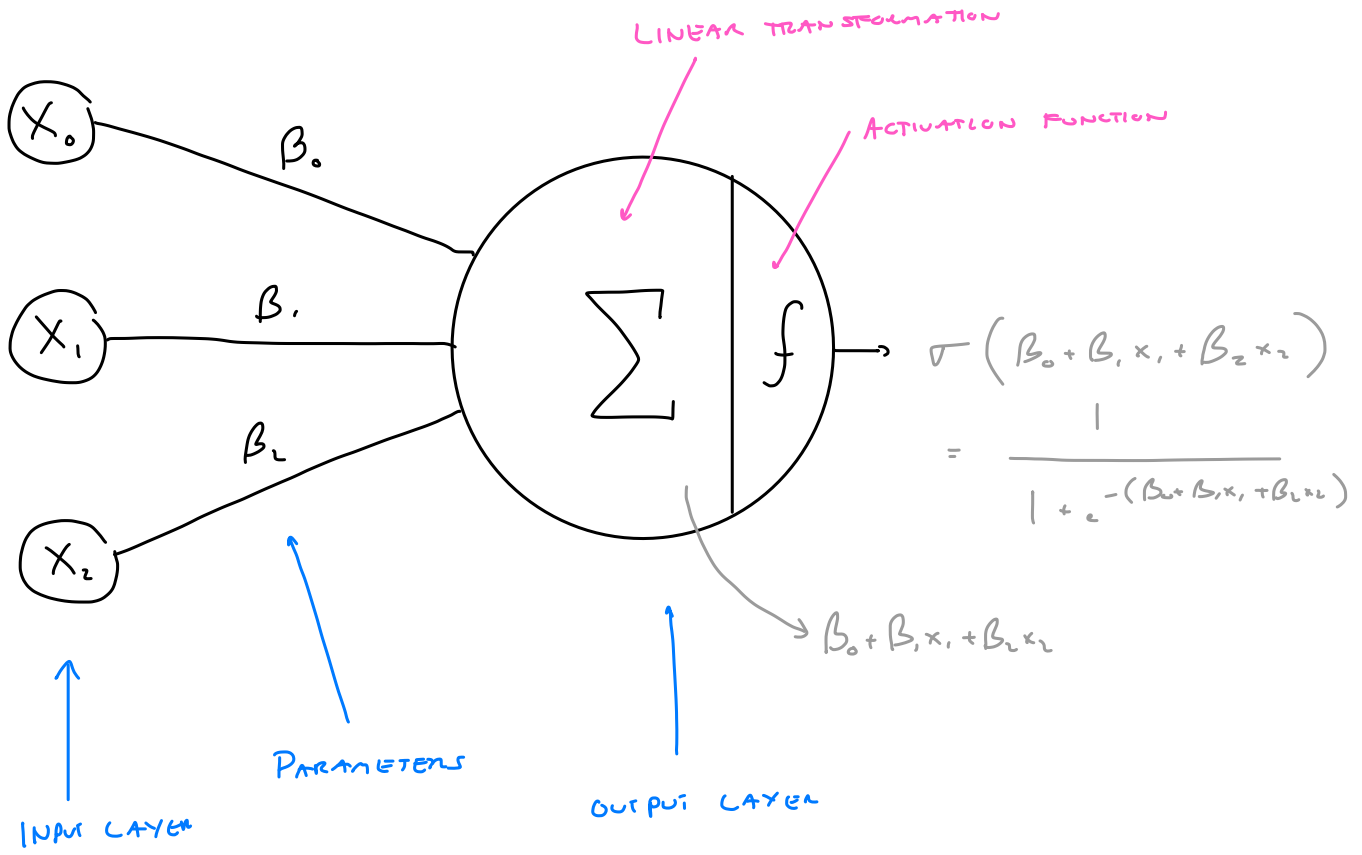
THREE PARAMETERS

TWO INPUT FEATURES

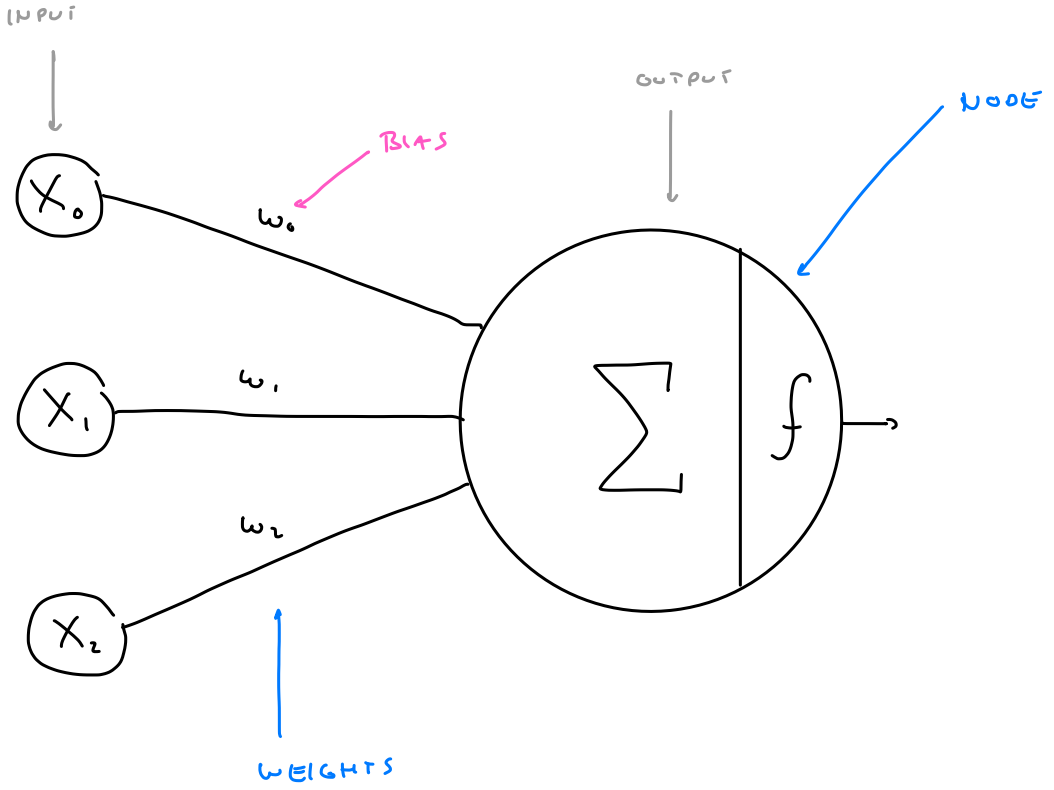
"LOGIT"

# LR AS A "NETWORK"

LET $X_0 = 1$



# NN NOMENCLATURE

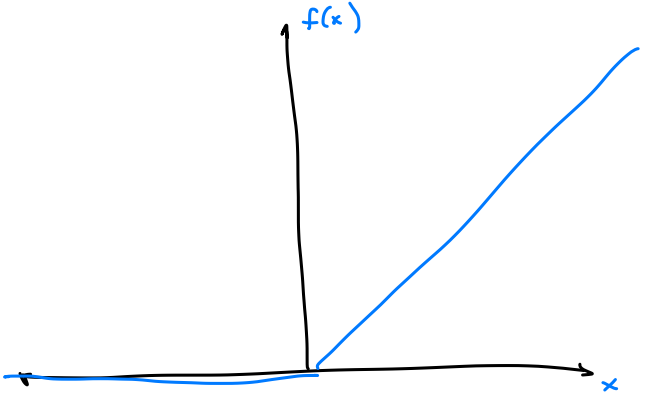


# ReLU

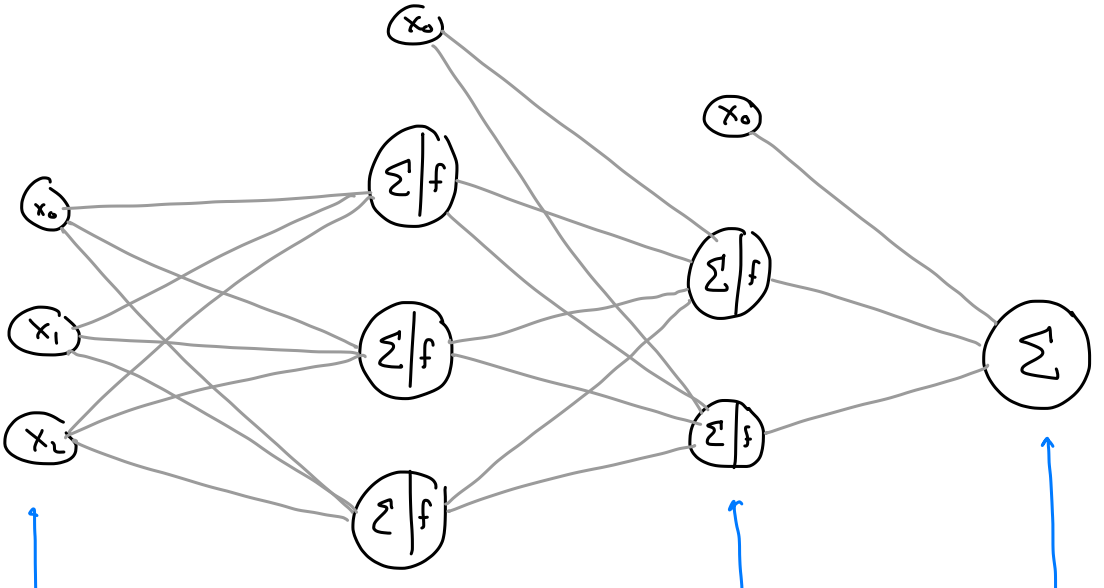
Rectified Linear Unit

Activation

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



# Going "Deep"




- Single output for binary
- One-per for multiclass

Input

Hidden
3 nodes
+ bias

Hidden
2 nodes
+ bias

Output

# 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
  - LEARNING RATE

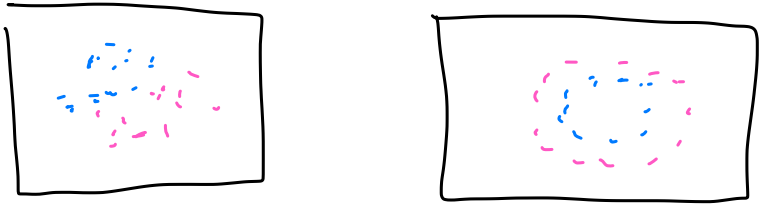
DEEP LEARNING = BIG NETWORKS

# How to train NNs?

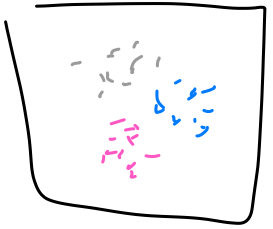
"How to learn weights from data?"

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

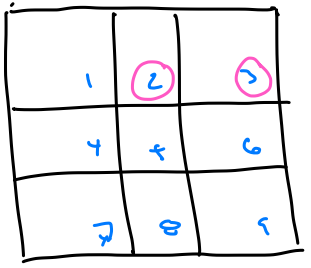
1. LEARN TOY BINARY DATA



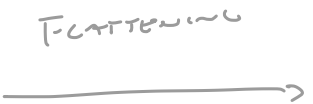
2. LEARN MULTICLASS DATA

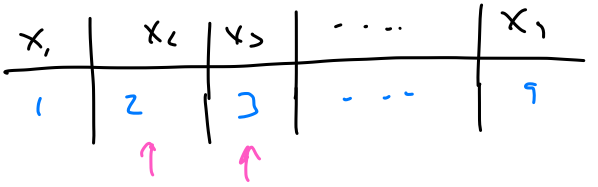


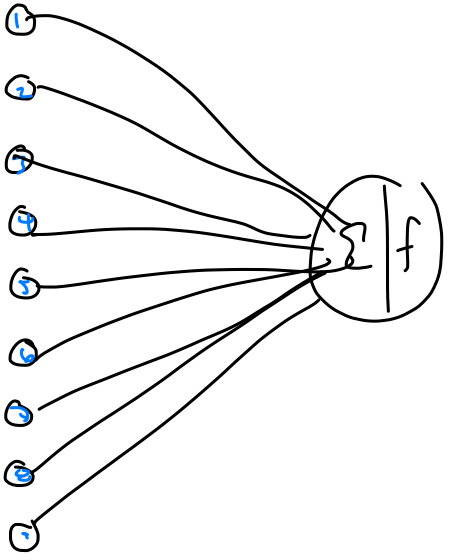
3. LEARN TO CLASSIFY IMAGES!




3x3



# Conv Net

## Input

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 5 | 6 | 7 |
| 1 | 0 | 2 |

3x3

## Kernel

| | |
|---|---|
| 1 | 0 |
| 0 | -1 |

WEIGHTS

2x2

## Output

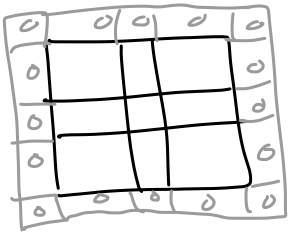
| | |
|---|---|
| -7 | -9 |
| 5 | 6 |

2x2

$1(1) + 0(2) + 0(5) - 1(6) =$

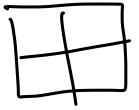
# Padding

Input



3x3

Kernel



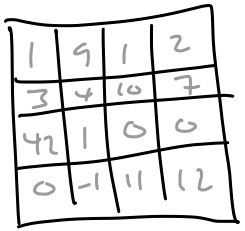
2x2

Output

4x4

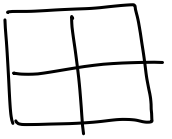
# Pooling — MAX / Average

Input




4x4

Kernel



2x2

Output

| | |
|---|---|
| 9 | 10 |
| 42 | 12 |