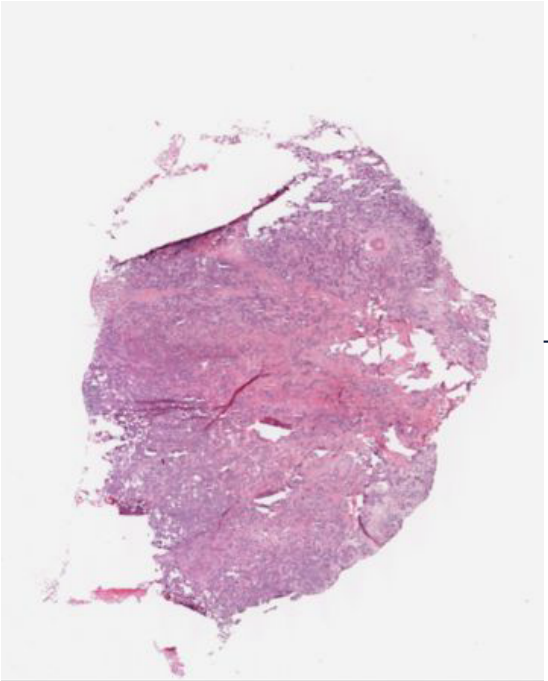
Neural Networks

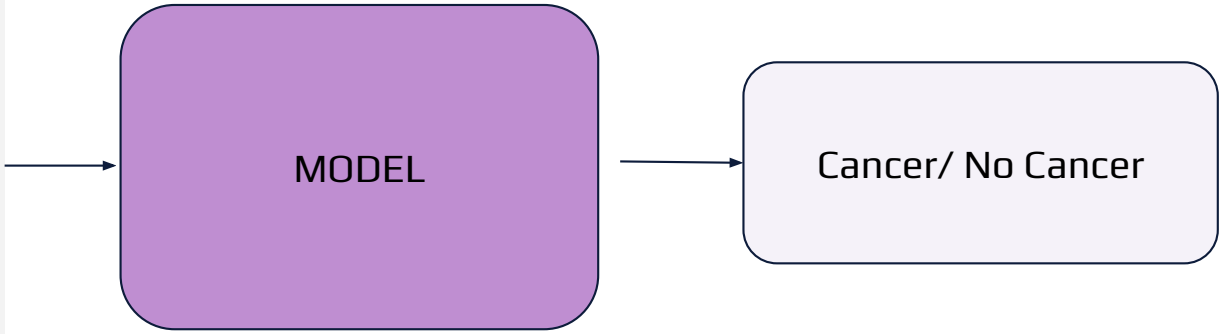
CPH 101A

Agenda

1. Motivation
2. Introduction to Neural Networks
3. Components
4. How training works

Goal: identify cancer slides





Pathology slides are **huge**

Satellite image of Manhattan

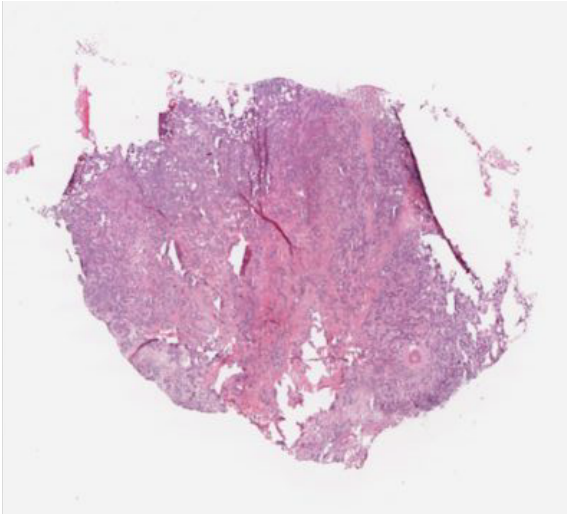
~30k pixels x 30k pixels

~9km x 9km

Zoom in

1 car ~ 15 pixels

Pathology slides are **huge**

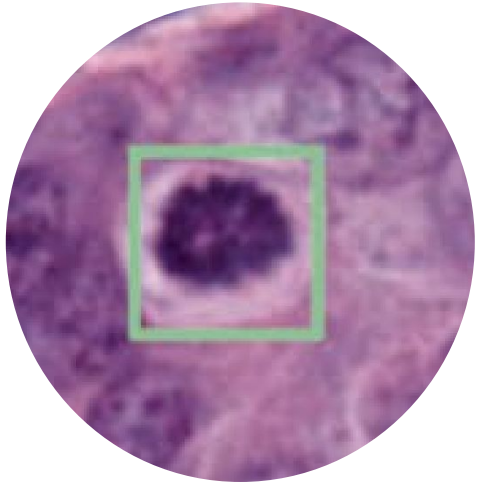


Pathology slide

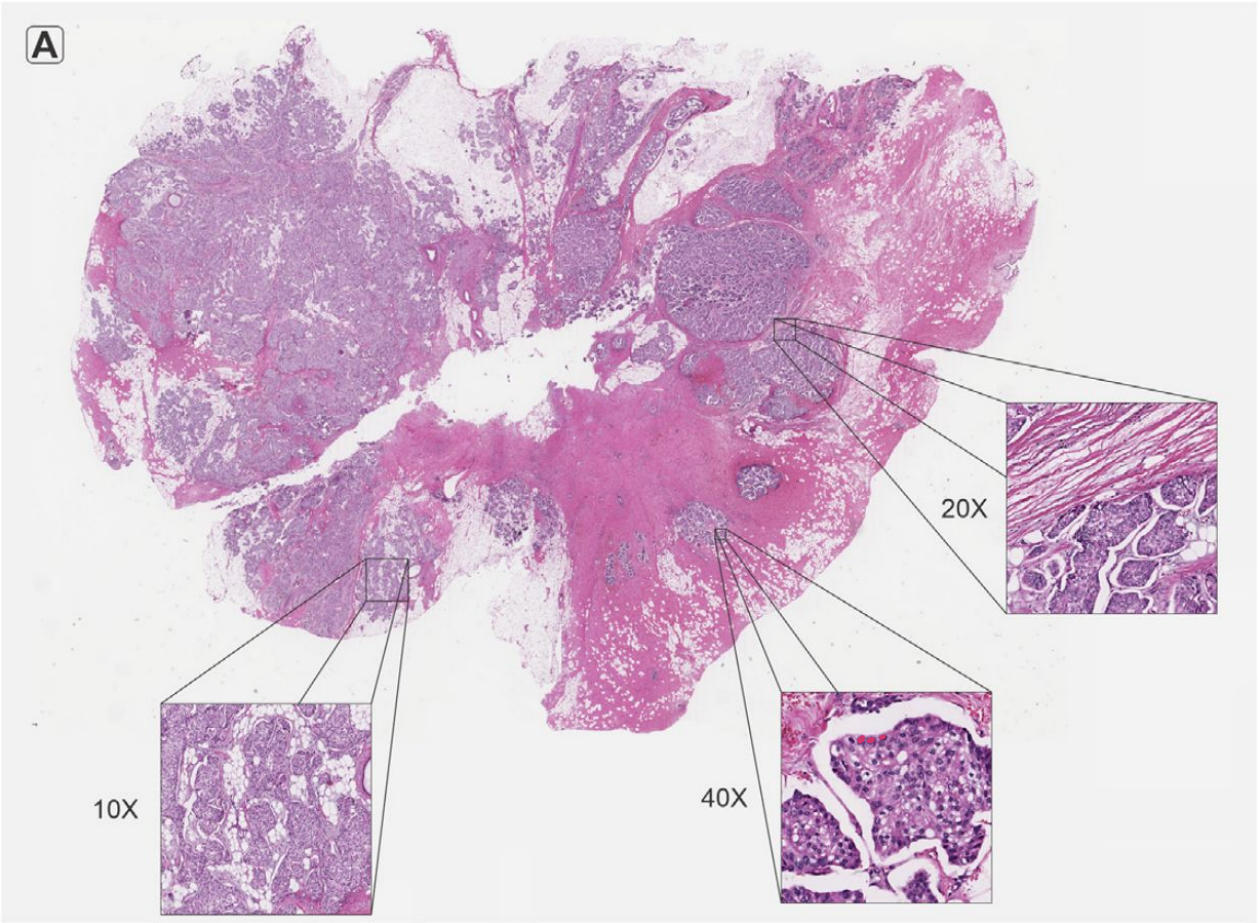
~30k pixels x 30k pixels

~1.5 cm x 1.5 cm (0.6 inch)

Zoom in

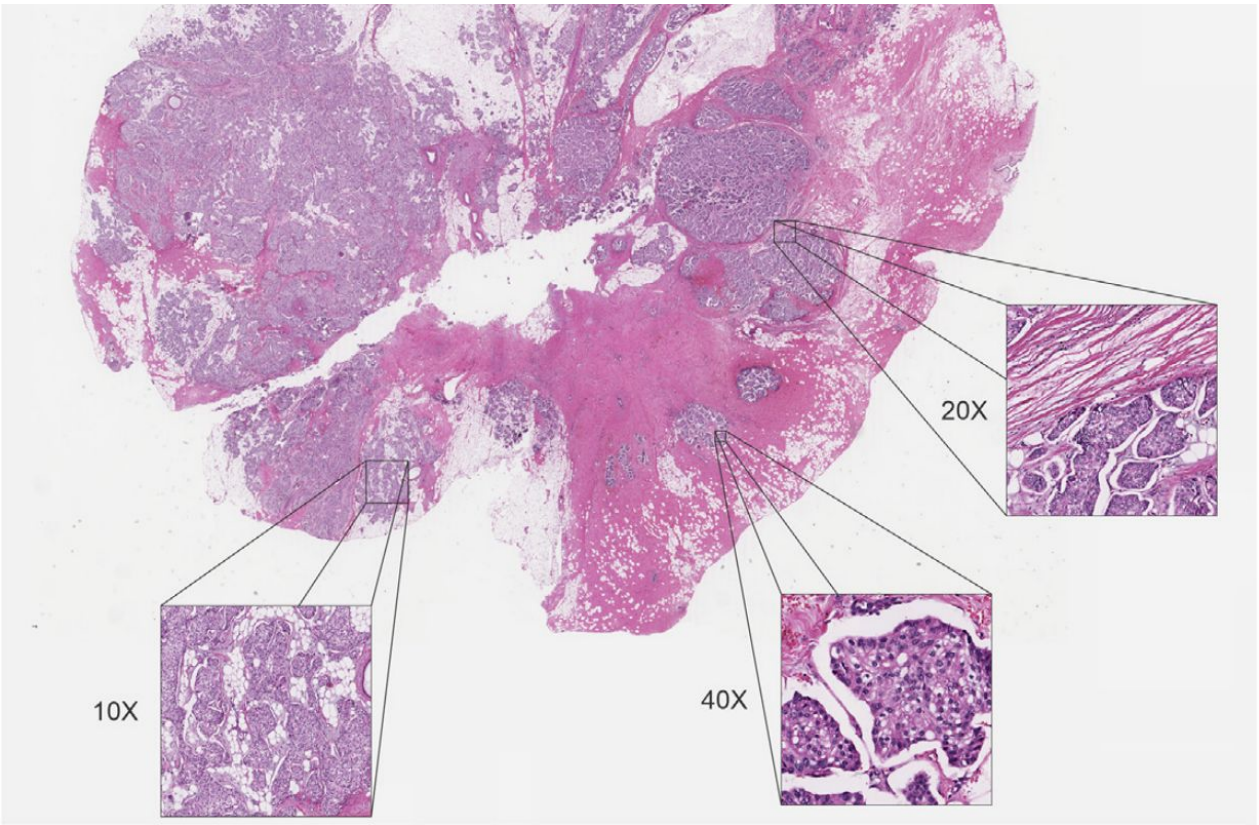


1 cell ~ 15 pixels



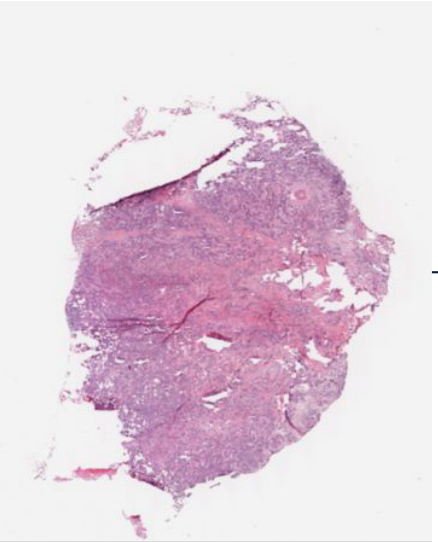
An example of breast H&E WSI from TCGA-BRCA dataset: (A) Illustration of 3 tile images in 10×, 20×, and 40× magnifications in a WSI.
From: Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review

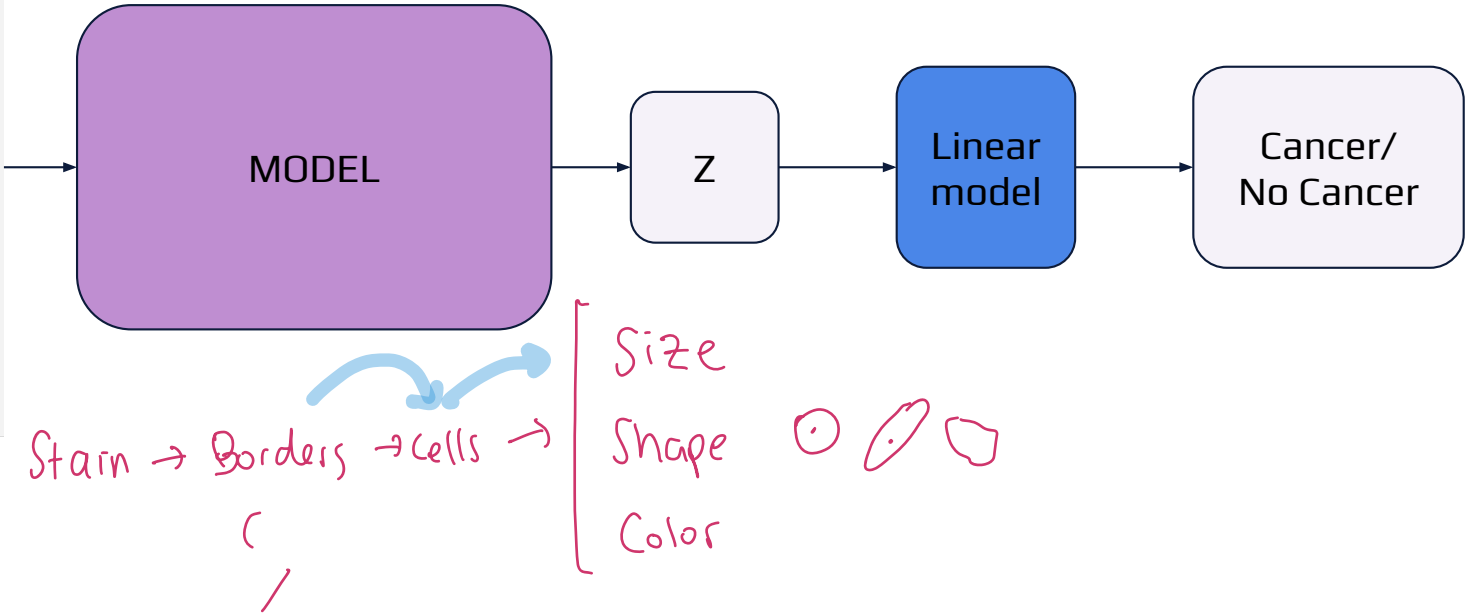
Complex relationship between pixels and outcome

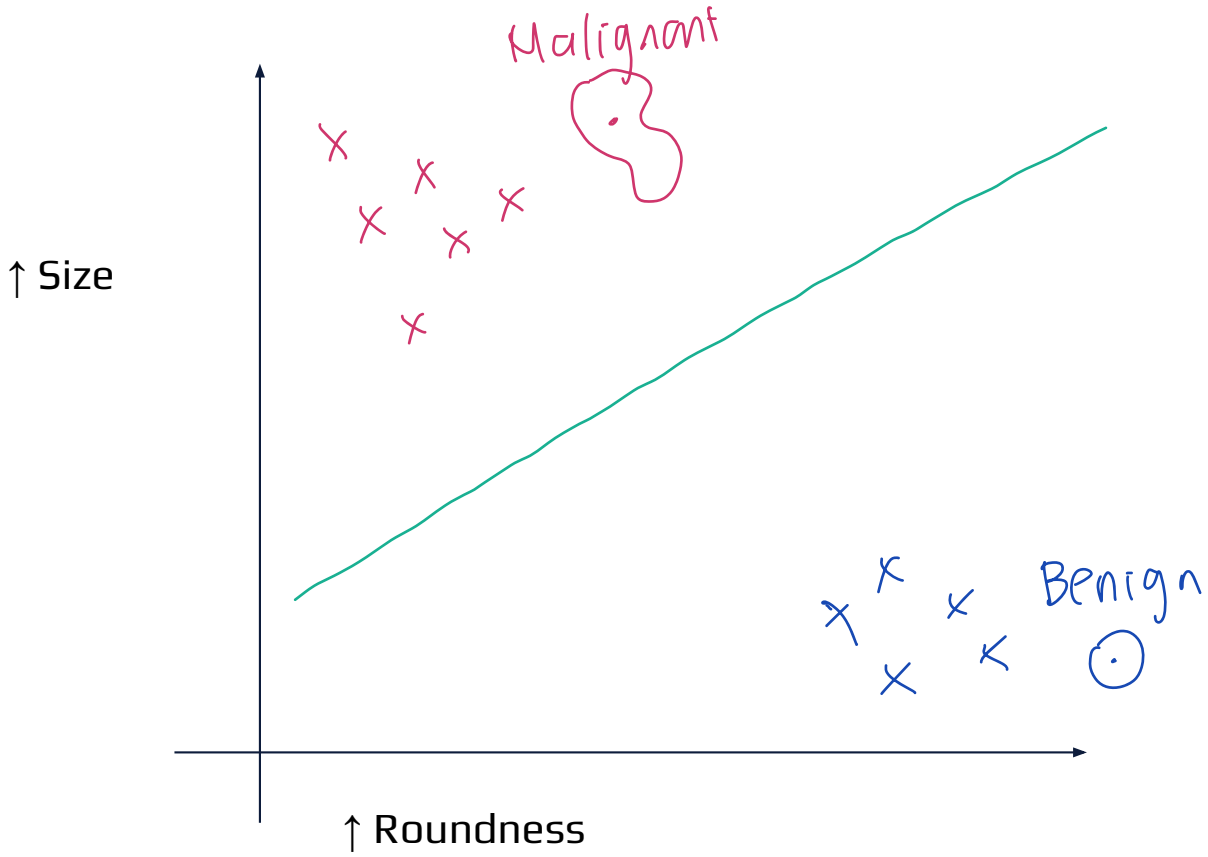


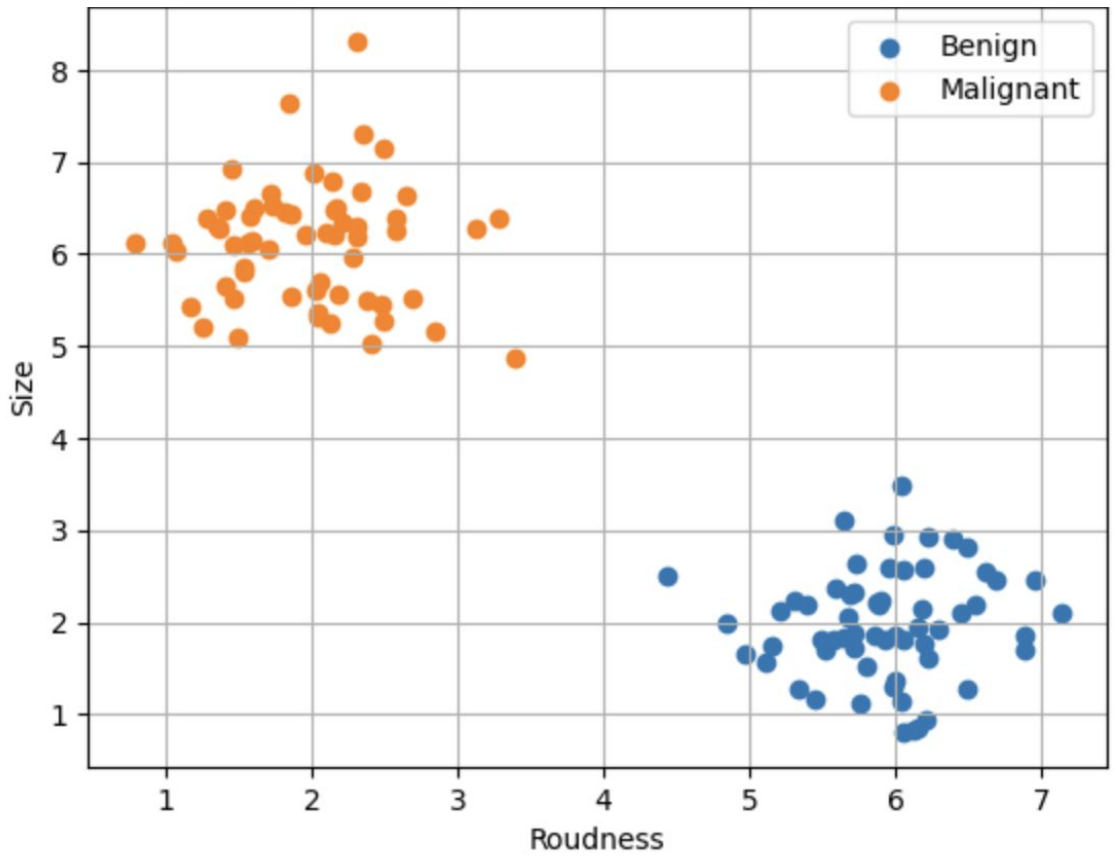
An example of breast H&E WSI from TCGA-BRCA dataset: (A) Illustration of 3 tile images in 10×, 20×, and 40× magnifications in a WSI.
From: Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review

Intermediate features might work

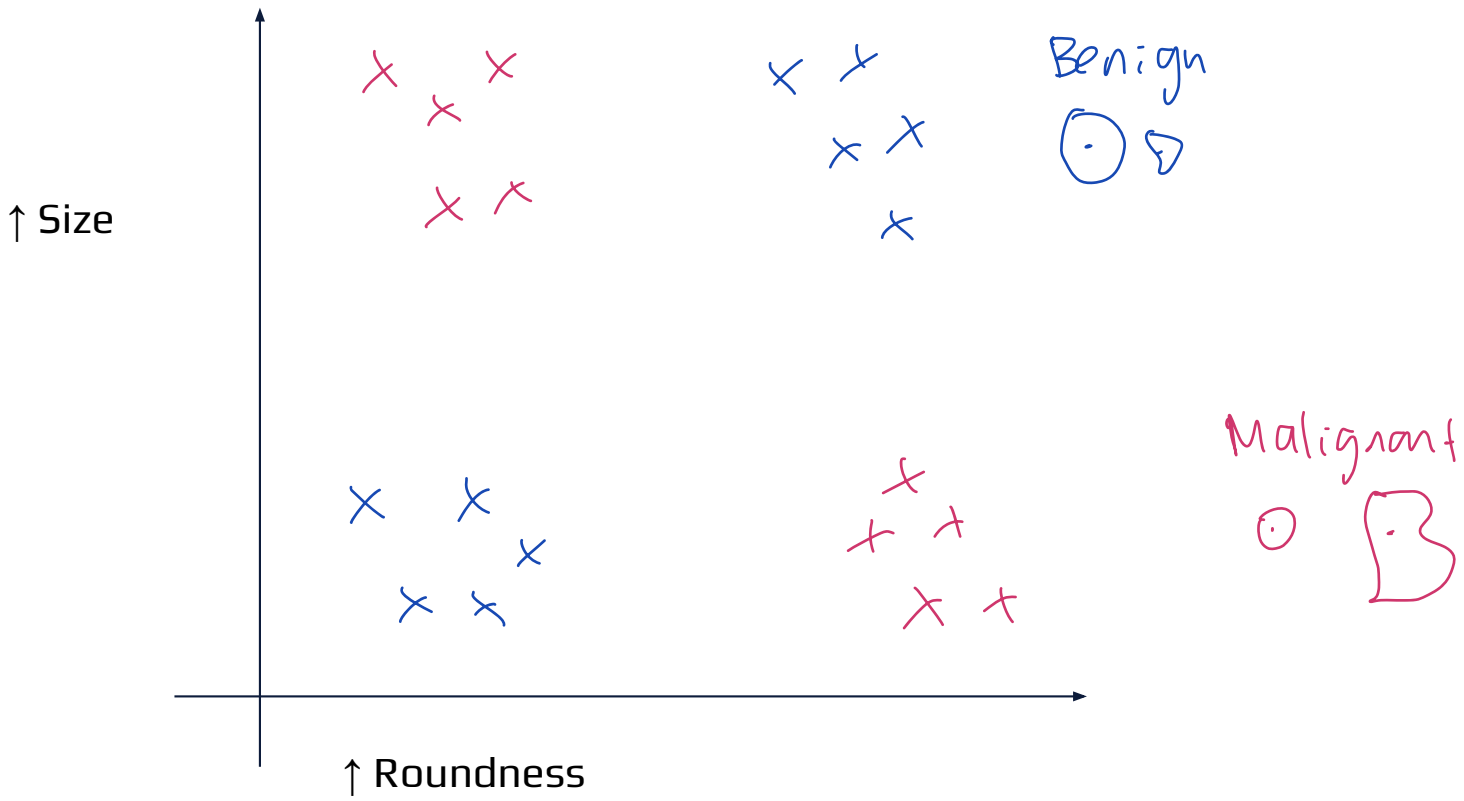


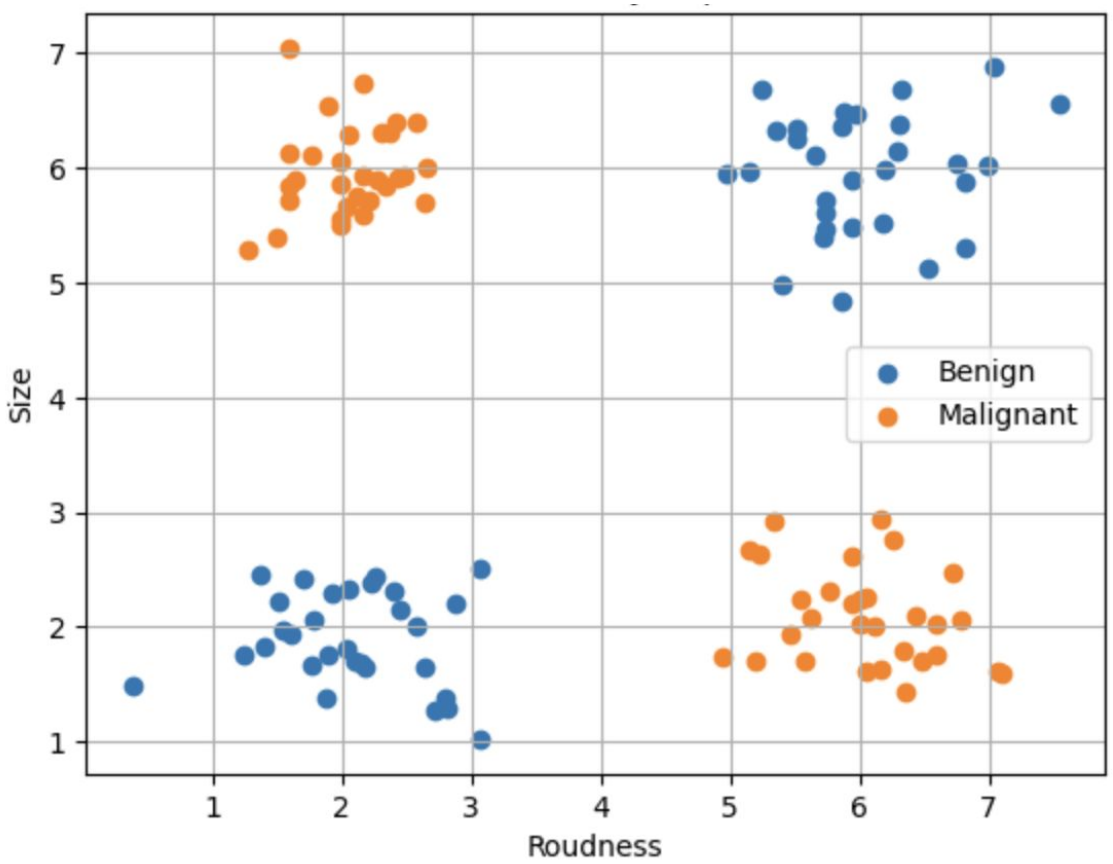



Malignant
↑ Size
Benign
↑ Roundness

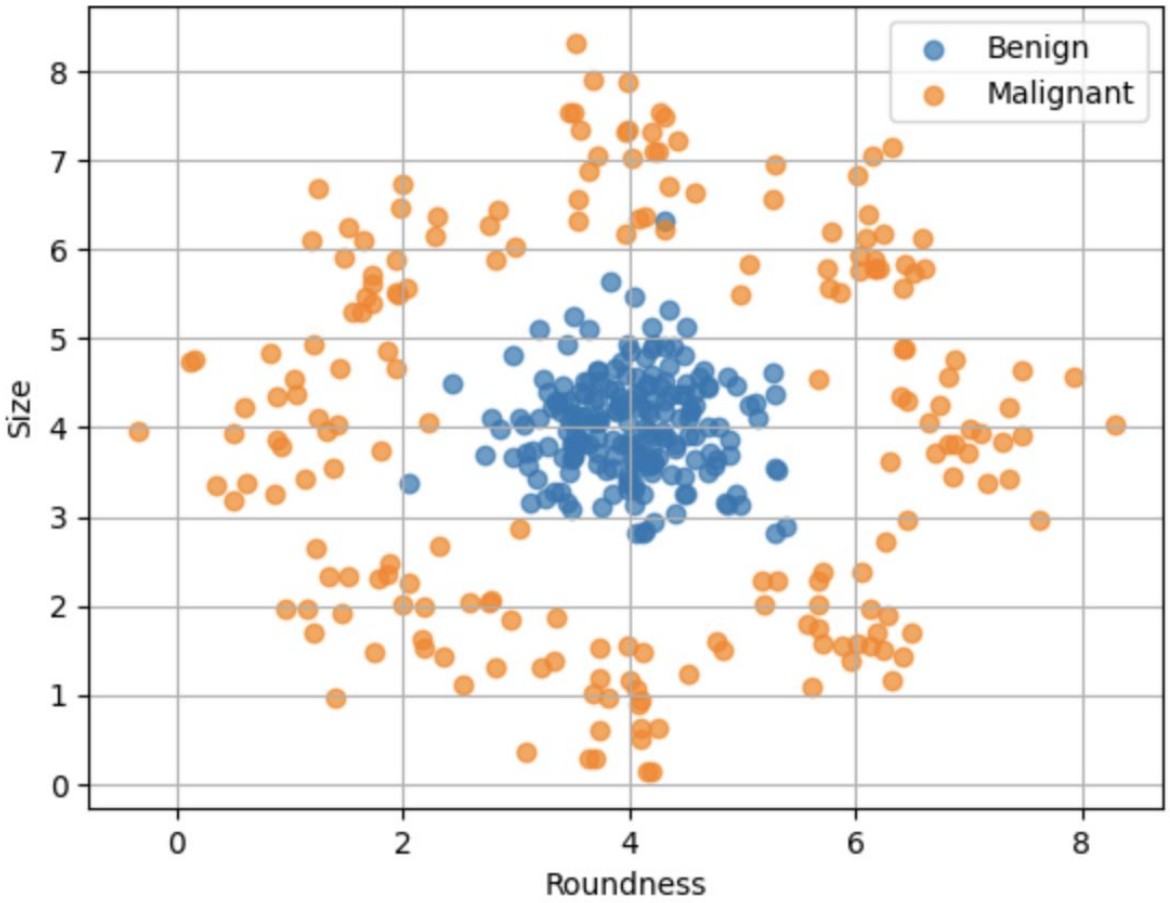


Intermediate features that are linearly separable


↑ Size
↑ Roundness
Benign
Malignant



Intermediate features that are NOT linearly separable



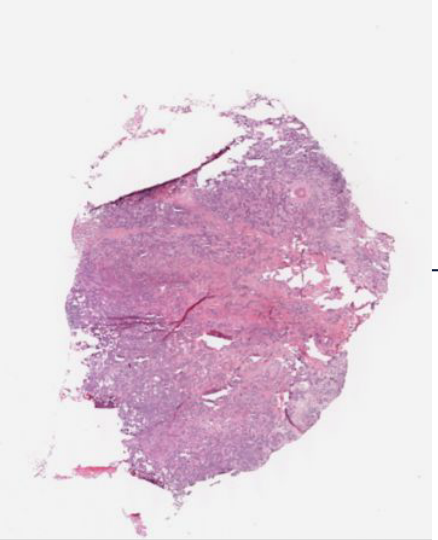
Intermediate features that are NOT linearly separable

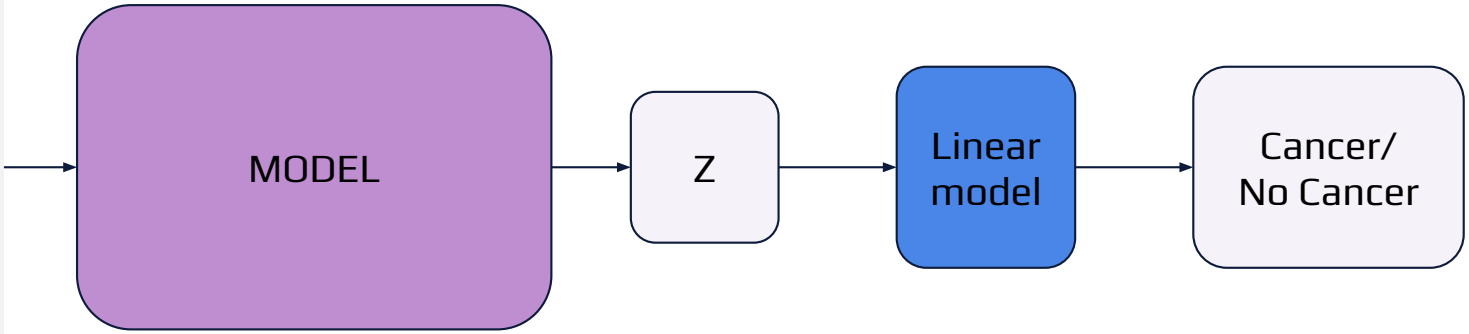
Problems with hand-crafted features:

Problems with hand-crafted features:

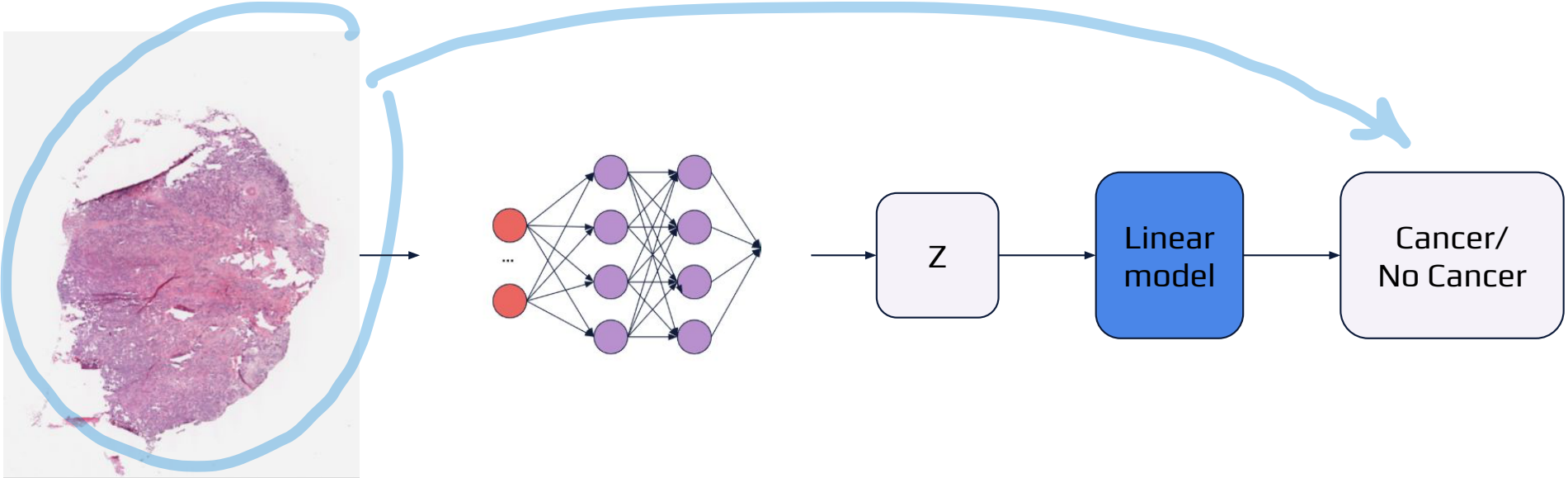
- Designing features is hard
- Missing out on other more complex interactions
- They don’t always work/progress stagnates
- Errors compound

Intermediate features might work:
Can we optimize the stages jointly?

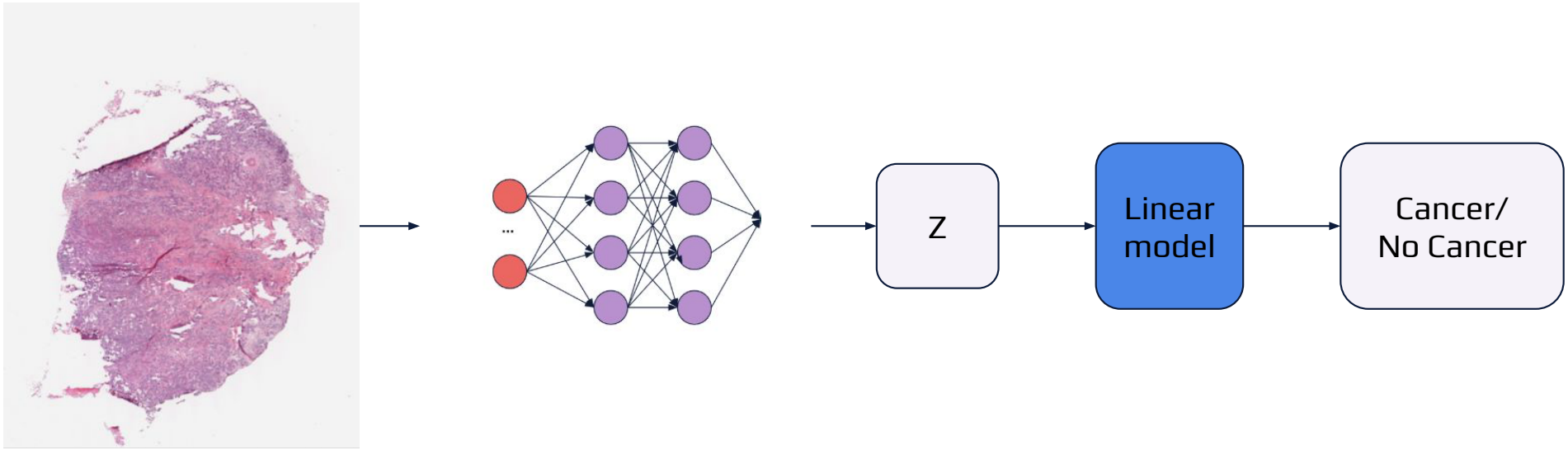




Intermediate features might work:
Can we learn them from scratch?

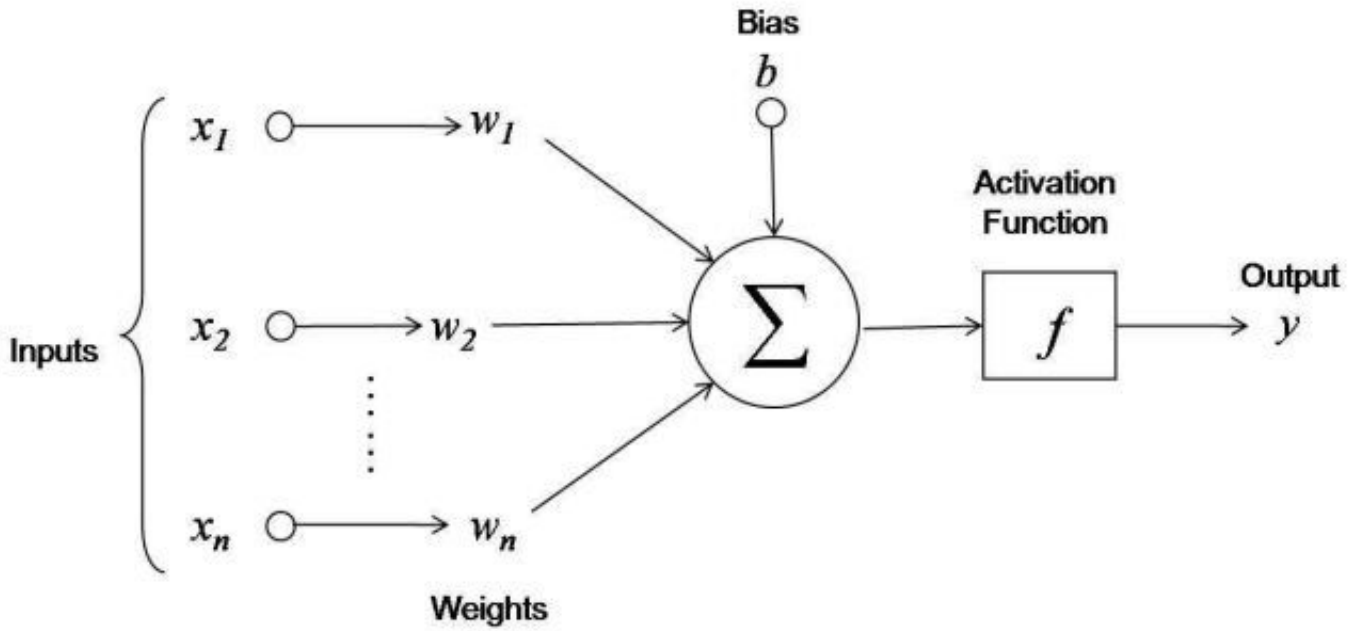


Intermediate features might work:
Can we learn them from scratch?

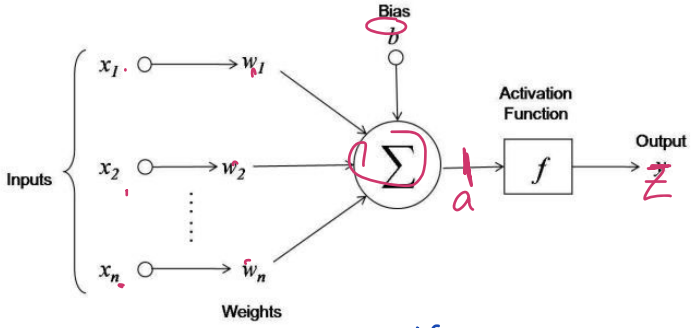


Trade-offs: explainability + more data needed to train

Neuron


Bias
b
Activation
Function
Output
y
Inputs
x_1
x_2
x_n
w_1
w_2
w_n
Σ
f
Weights

Neuron



x input

$x \in \mathbb{R}^n$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$\theta \in \mathbb{R}^n$

weights

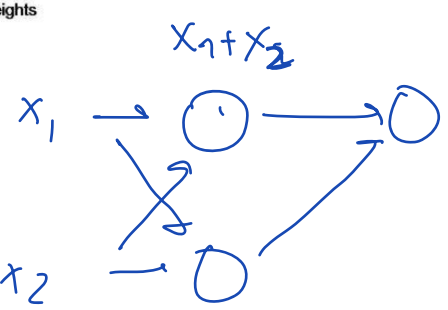
$$x_1 w_1 + x_2 w_2 + \dots + x_n w_n + b = a$$

$b \in \mathbb{R}$

bias

$$a = \theta x + b$$

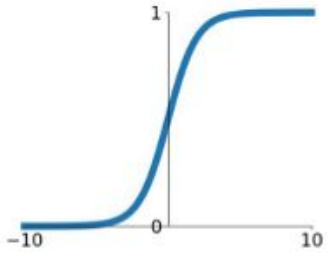
$$z = f(a)$$



Common activation functions

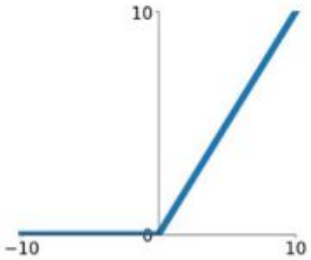
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

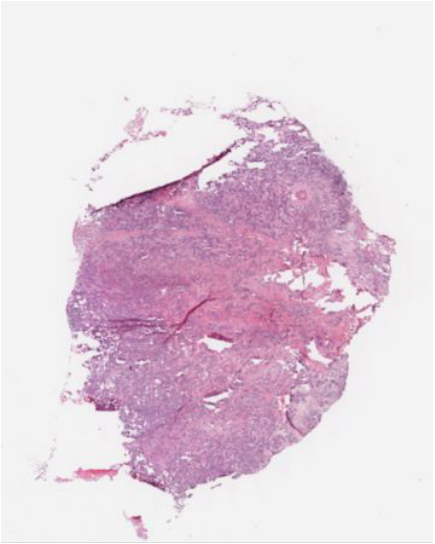


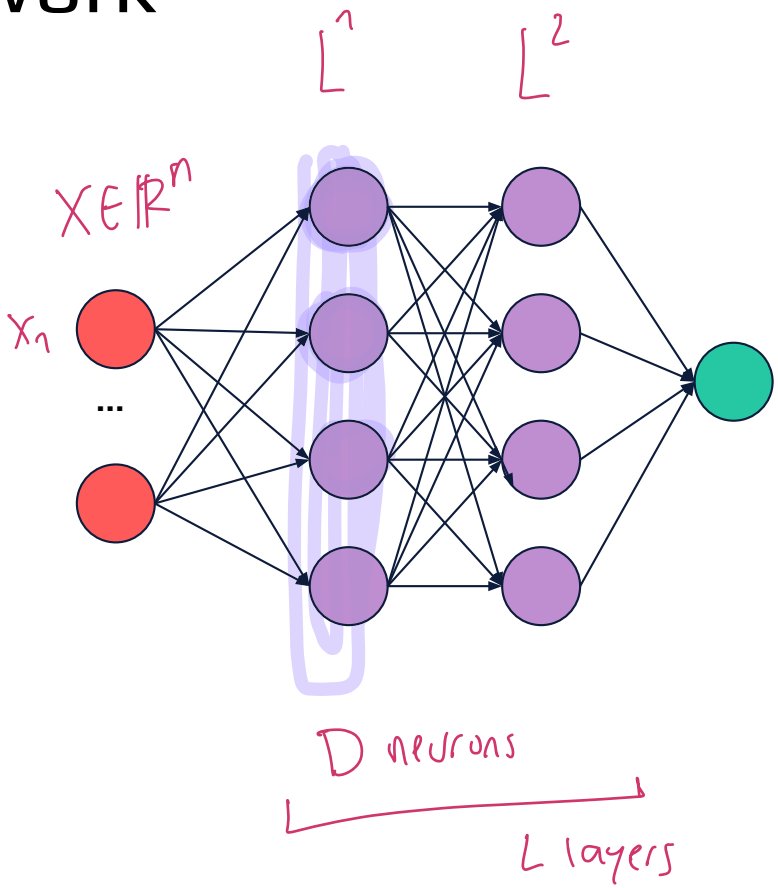
ReLU

$$\max(0, x)$$

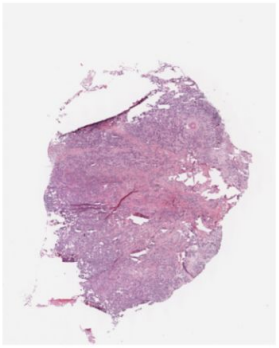


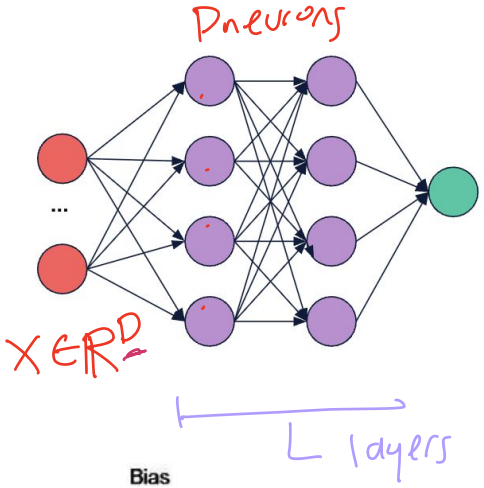
Neural Network





How many parameters we need to learn?





ONE NEURON $\theta \in \mathbb{R}^D$ weights, $\eta = D$

$b \in \mathbb{R}$ bias

$a = \theta x + b$

$z = f(a)$

Layer 1) Neuron $D+1$

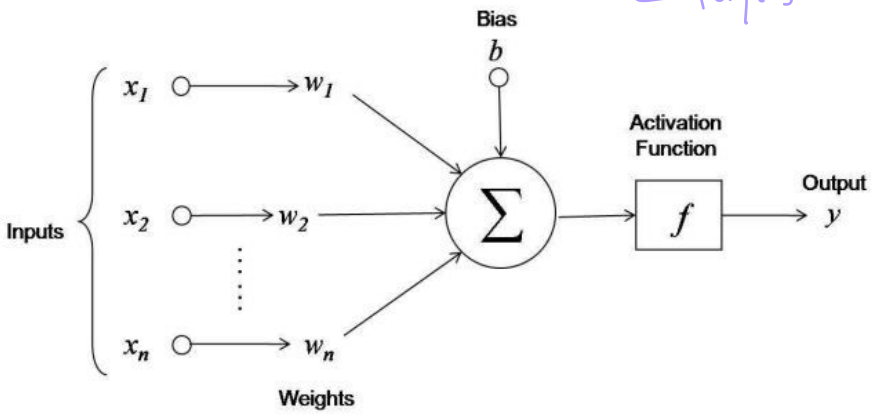
$\vdots$ $D+1$

$\vdots$

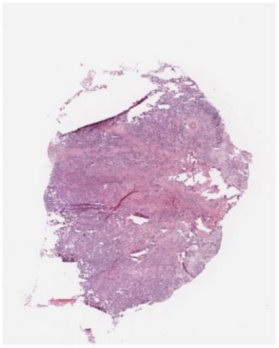
$D(D+1) = D^2 + D$

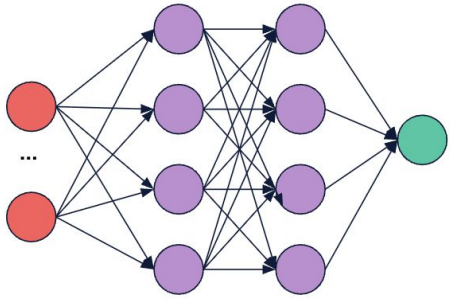
$\#: L(D^2+D) = LD^2 + LD$

$O(LD^2)$



How many multiplications we need to make?





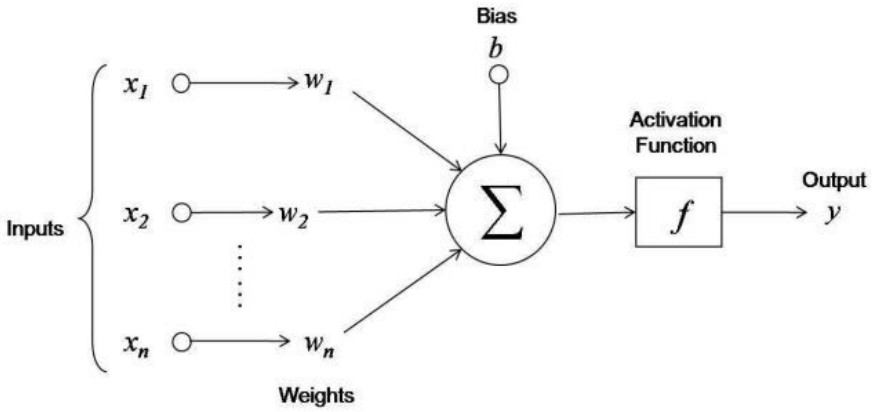
ONE NEURON $\theta \in \mathbb{R}^D$ weights $\eta = D$

$b \in \mathbb{R}$ bias

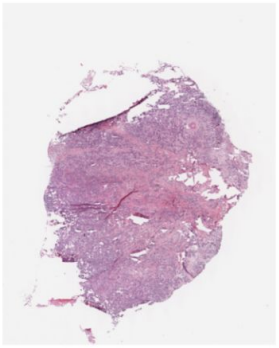
$a = \theta x + b$

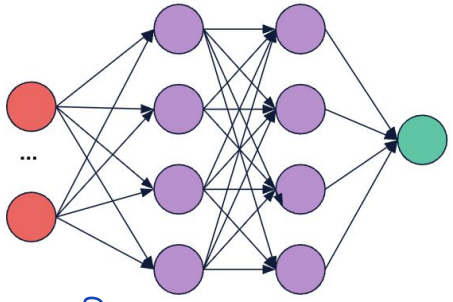
$z = f(a)$

$L D^2$



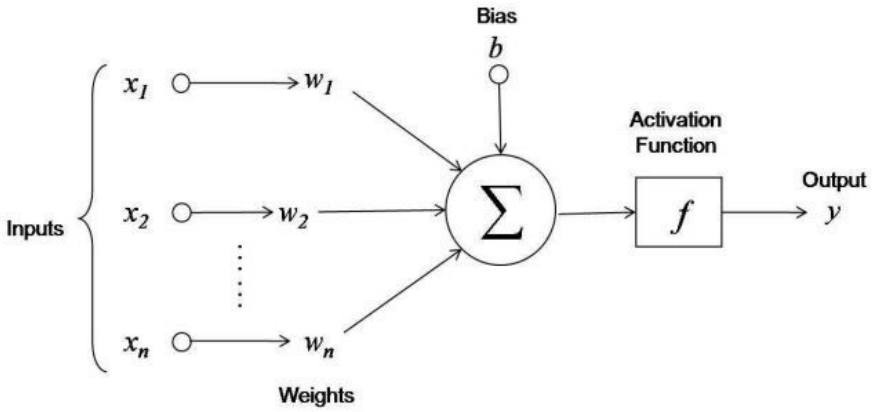
How much memory we need for inference?



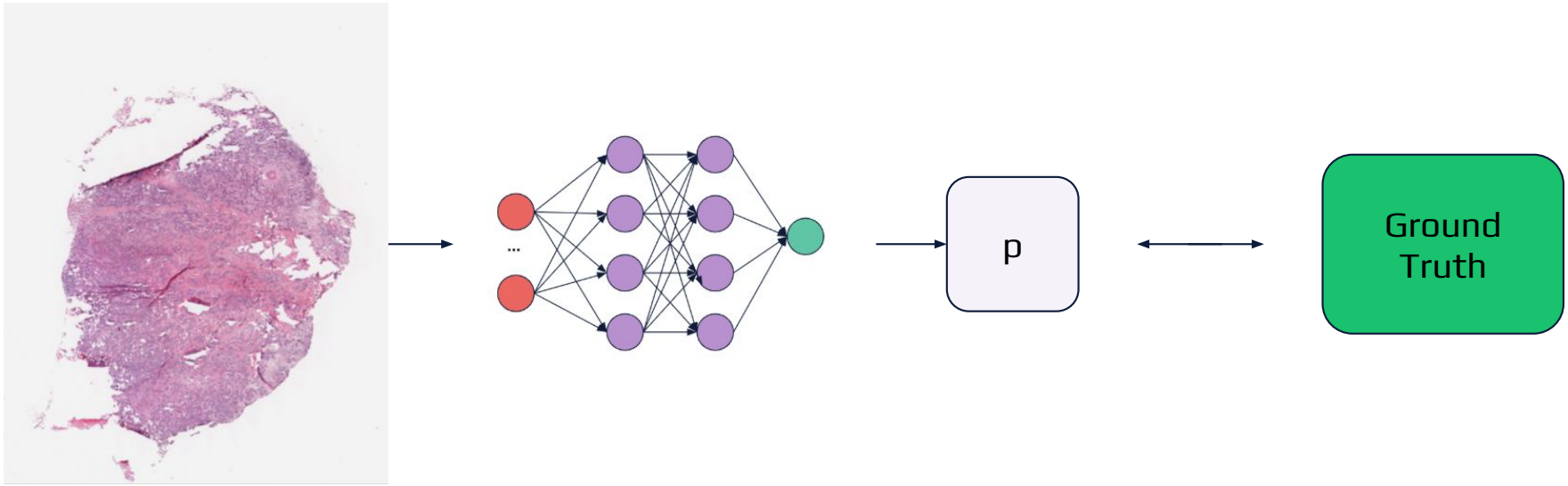


$x \in \mathbb{R}^D$

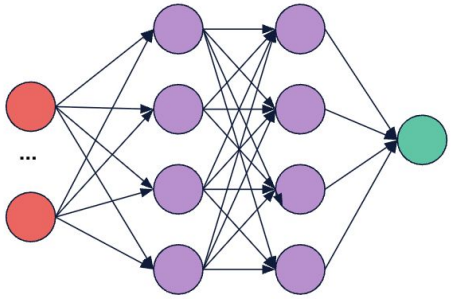
$O(D)$



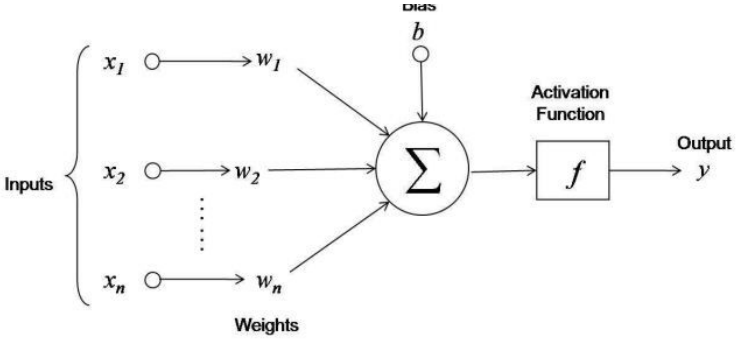
How do we train a Neural Network?



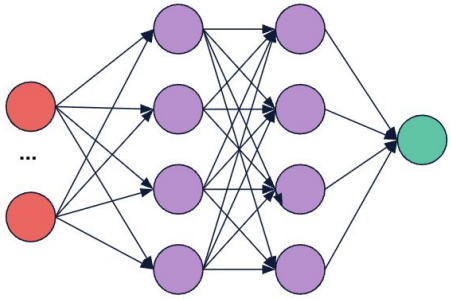
How do we train a Neural Network?

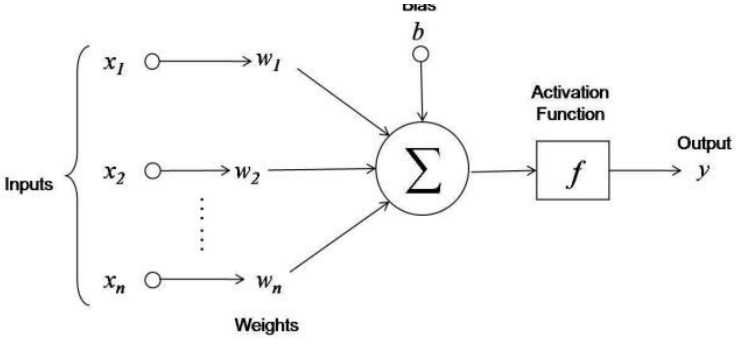


1) Loss



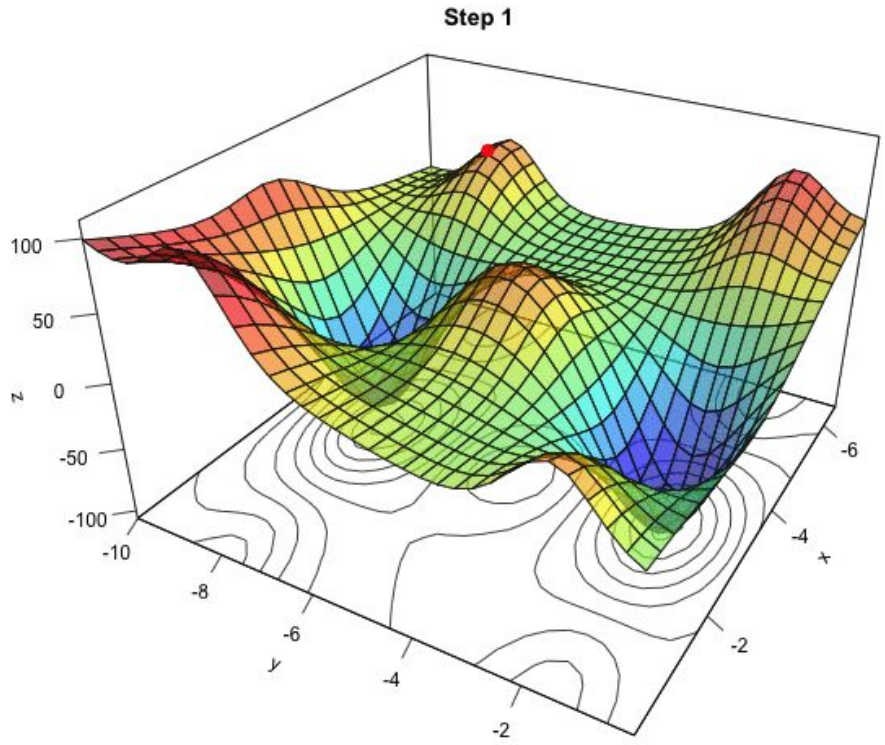
How do we train a Neural Network?



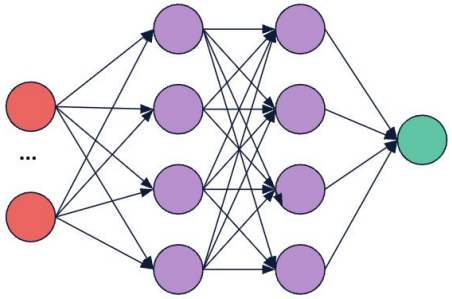


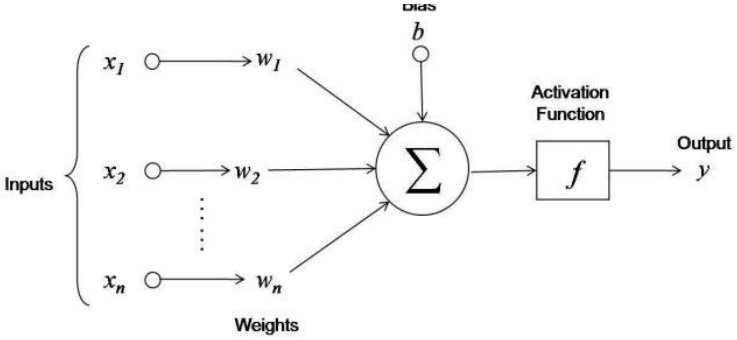
2) Gradient Descent

Gradient Descent



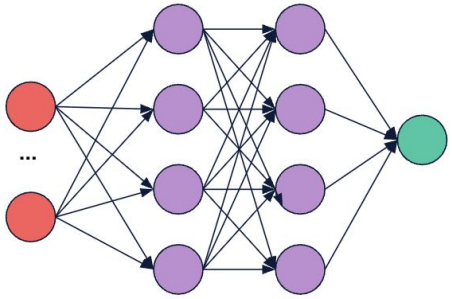
Reminder: Chain Rule



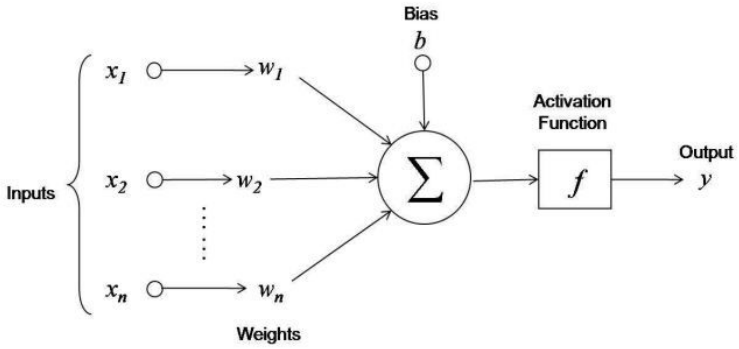


Gradient Descent

How do we train a Neural Network?

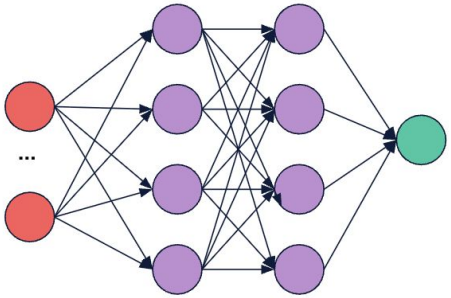


3) Backpropagation



Backpropagation

Backpropagation



https://eli.thegreenplace.net/2018/backpropagation-through-a-fully-connected-layer/