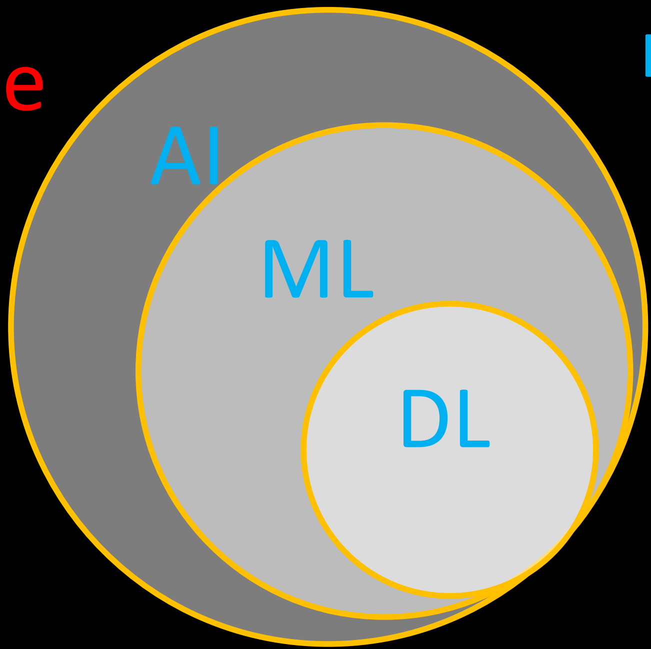
AI (Goal):
Artificial intelligence

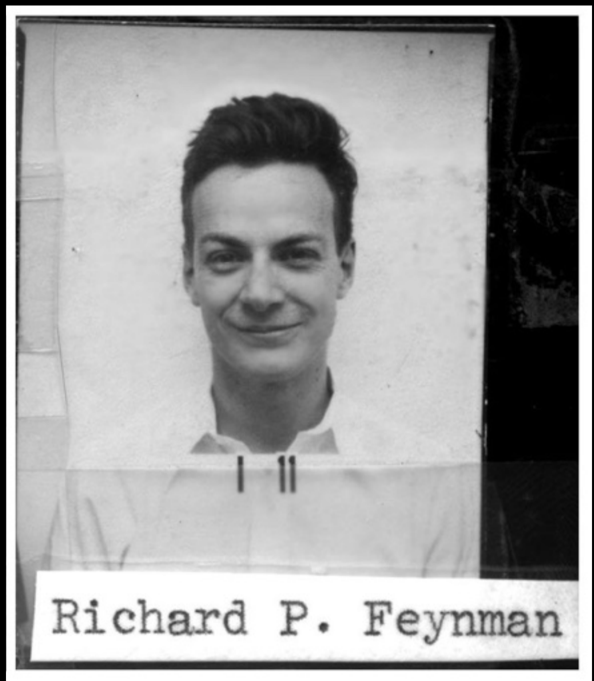


ML:
machine learning

DL:
deep learning

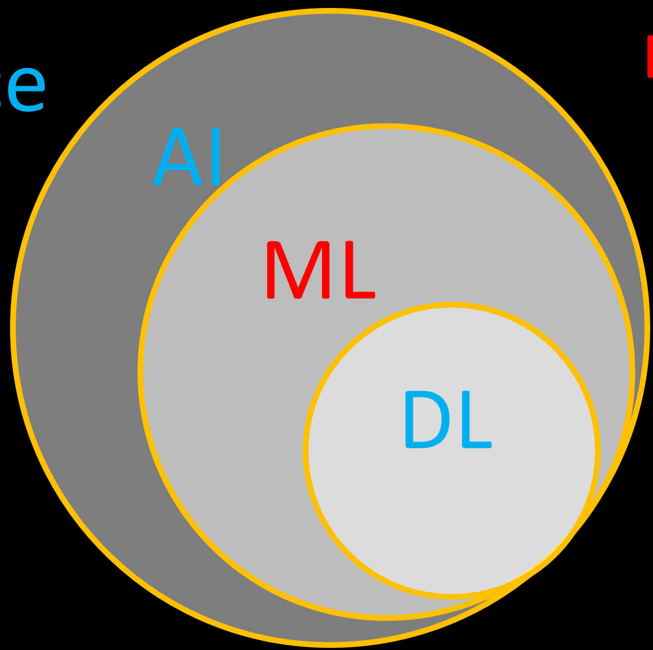
1. "human-level" intelligence has a diverse definition

2. AI may reach human level intelligence, but probably are not doing it in a "humane" way



AI (Goal):
Artificial intelligence

ML:
machine learning



DL:
deep learning

ML: machine learning

(collection of methods) to develop algorithms automatically by learning from data

Option 1: Manual-Crafted Rule Based

Truck

Wheels > 4

Truck

Height > 2m

Truck?

Color = red

Truck = (Wheels > 4)
and / or
(Height > 2m)

A Good Rule

Truck = (Color == red)

A Bad Rule

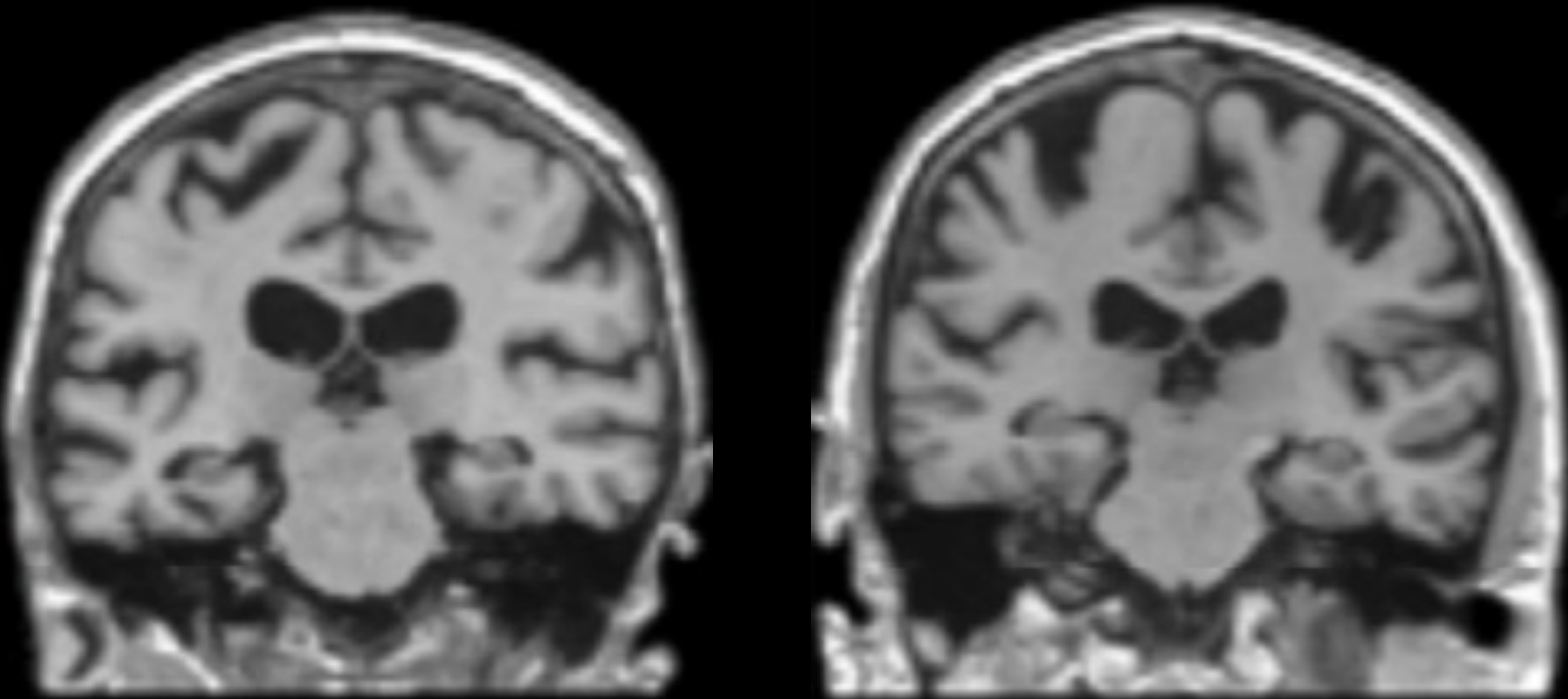
Hand crafted (explicit rules)

How to do define a function to detect tires?

get_circle

get_circle + is_tire
(black rims larger than….)

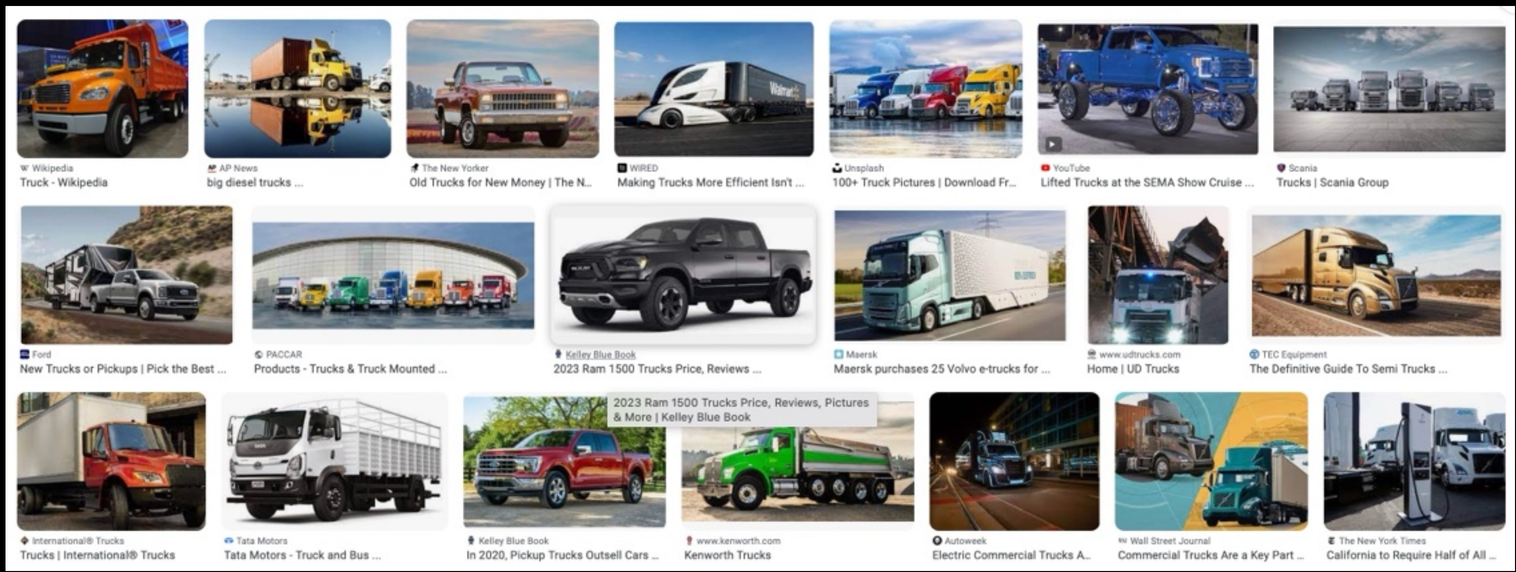
get_circle + is_tire +
tires_belong_to_the_same_car….



Complicated, real-life problems can’t simply be resolved by rules

Option 2: let the model to learn from data: Machine Learning

$f($ INPUT $)$ ⟶ OUTPUT



Model + Data + ways for the model to improve itself

Machine learning:

Learning a model that don't have explicit rules, instead the model is derived implicitly from observations (data)

when,

Rules are hard to be implemented
&
We don't know the rules clearly.

Look for Function(**s**)

$$f(\text{INPUT}) \longrightarrow \text{OUTPUT}$$

Visuals, audios, text…

Labels, predictions, decisions…

Look for Function(**s**)

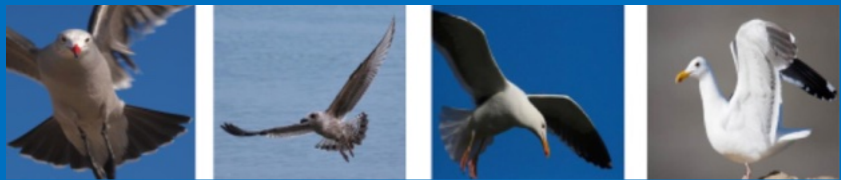
$f($ INPUT $)$ ⟶ OUTPUT

"Magic"

(turn out it's neural networks)

Machine Learning: Look for Function(**s**)

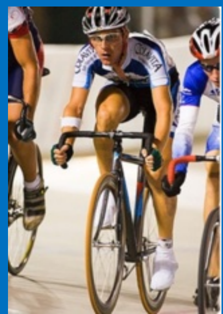
Classification

$f($  $)$ ⟶ Class (what kind of bird)

Detection

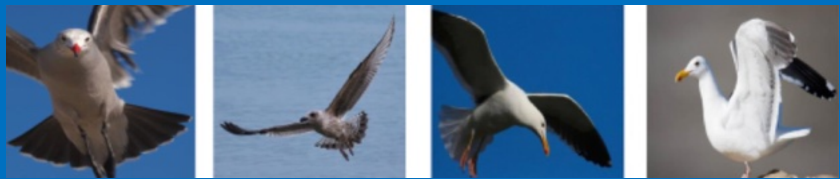
$f($  $)$ ⟶ Class (people lamp), Box location, Box size

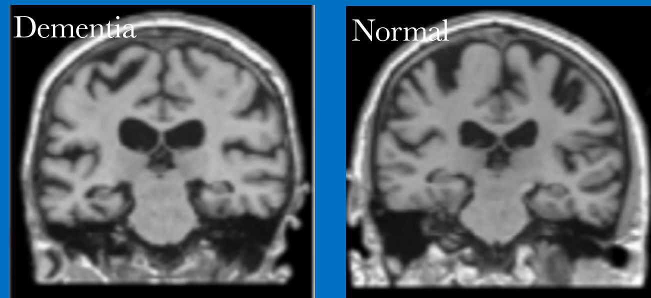
Segmentation

$f($  $)$ ⟶ 

Deep Learning in Medicine

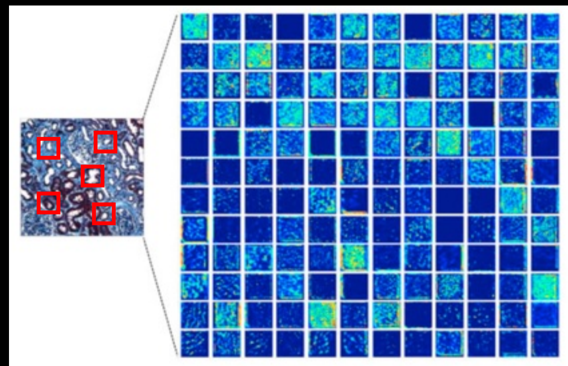
Classification





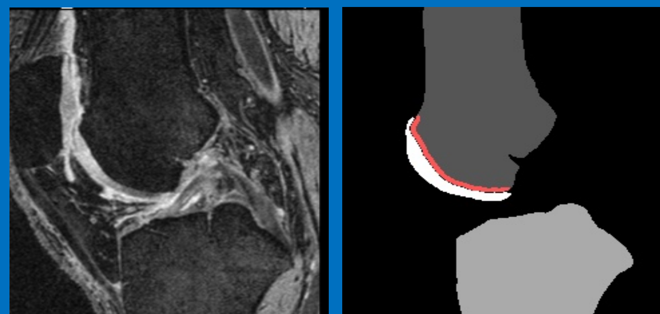
Dementia Prediction

Detection



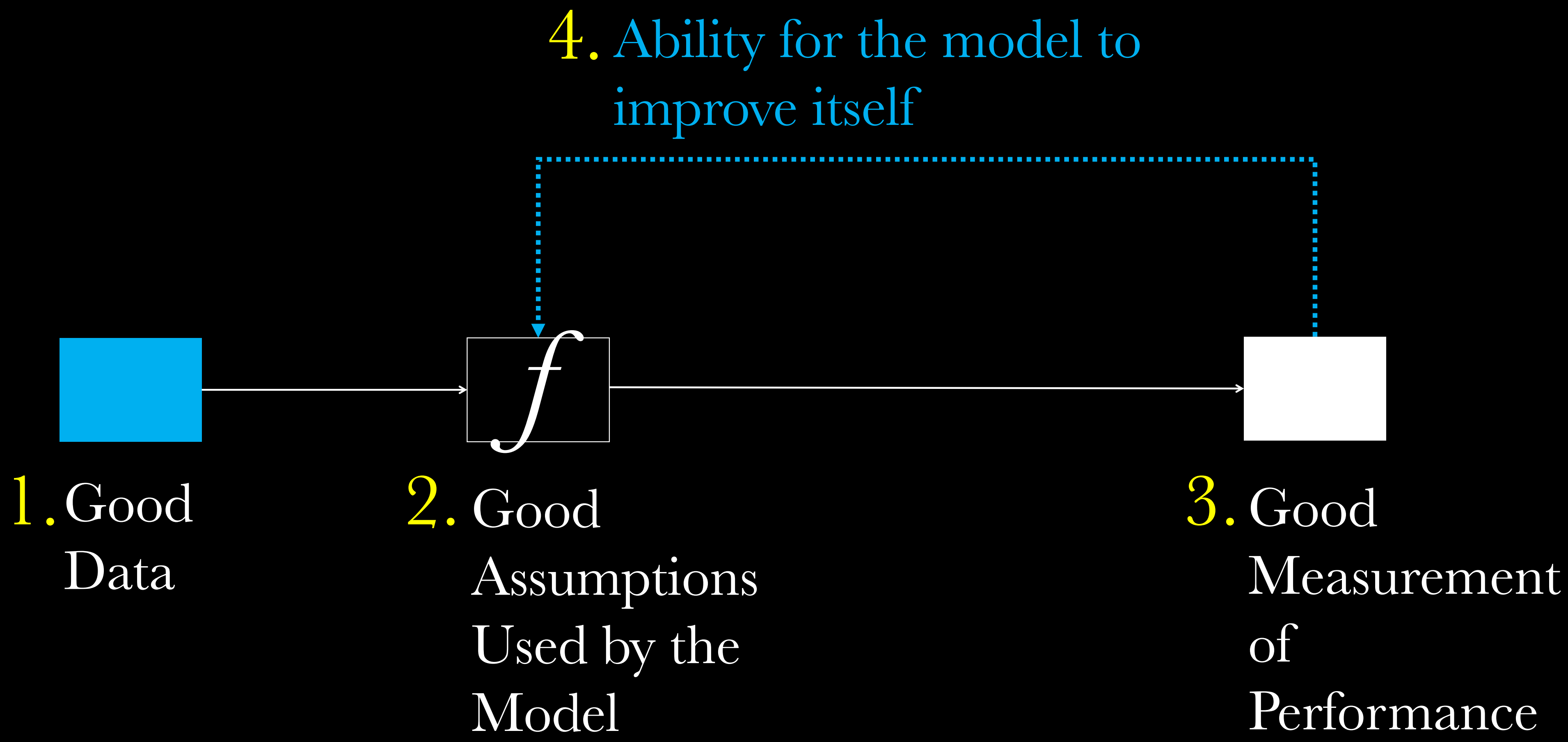
Hypocellularity Detection

Segmentation

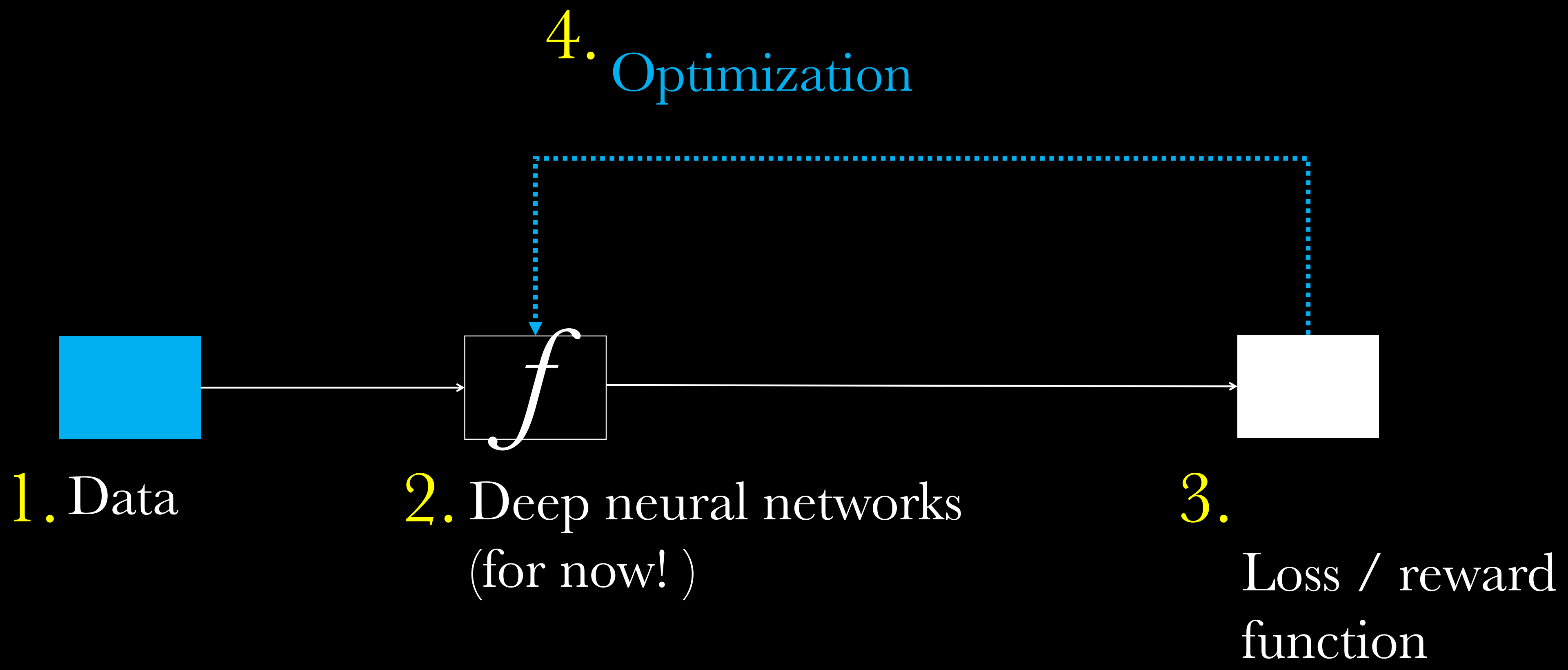


knee anatomical segmentation

Four Horseman of ML

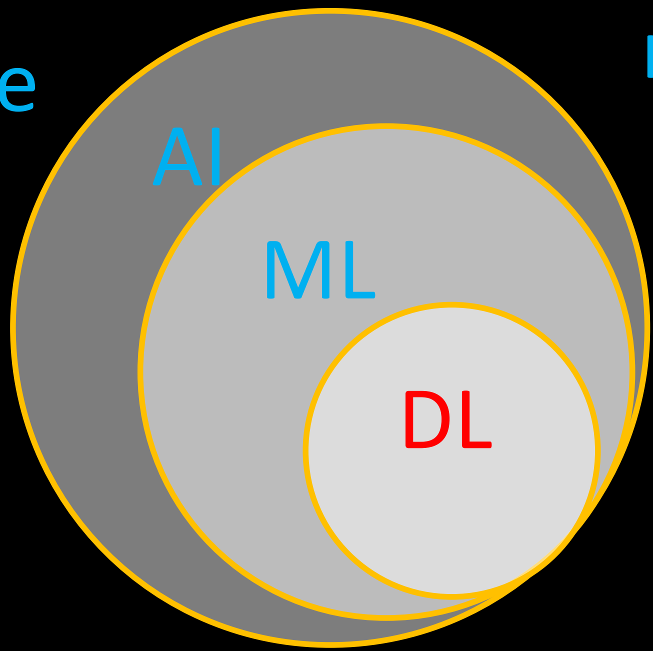


Four Horseman of ML

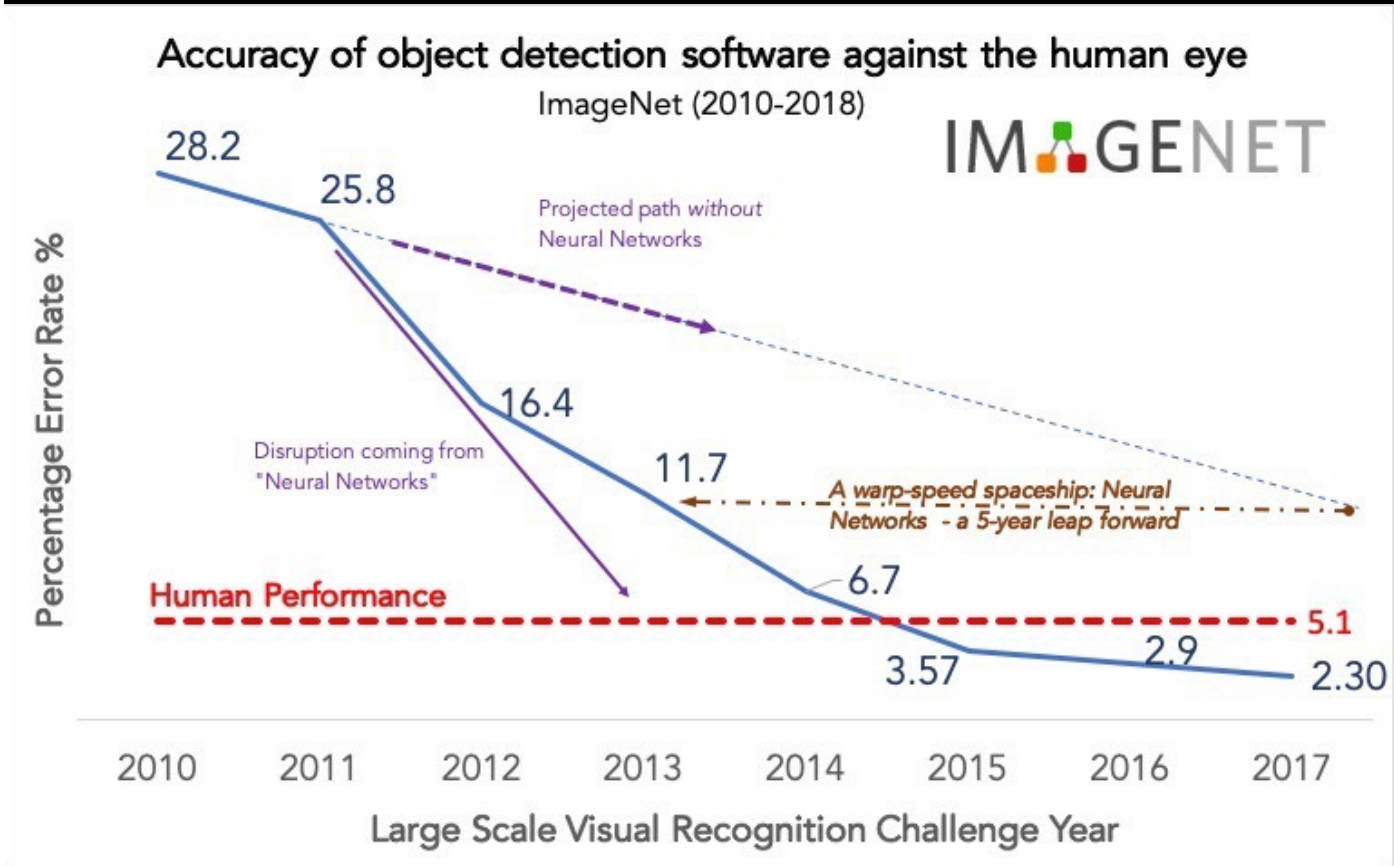


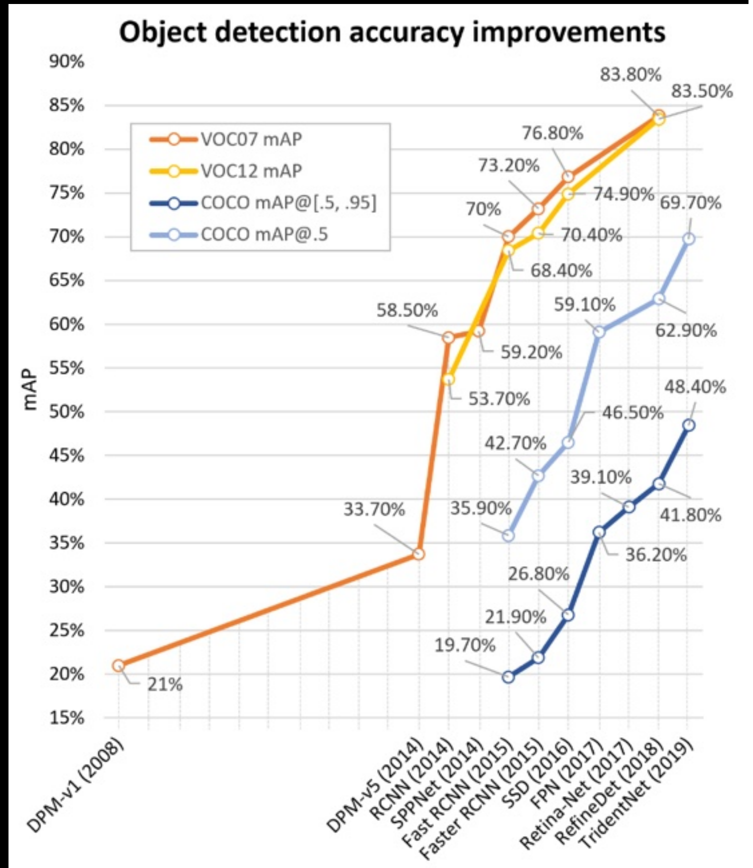
AI (Goal):
artificial intelligence

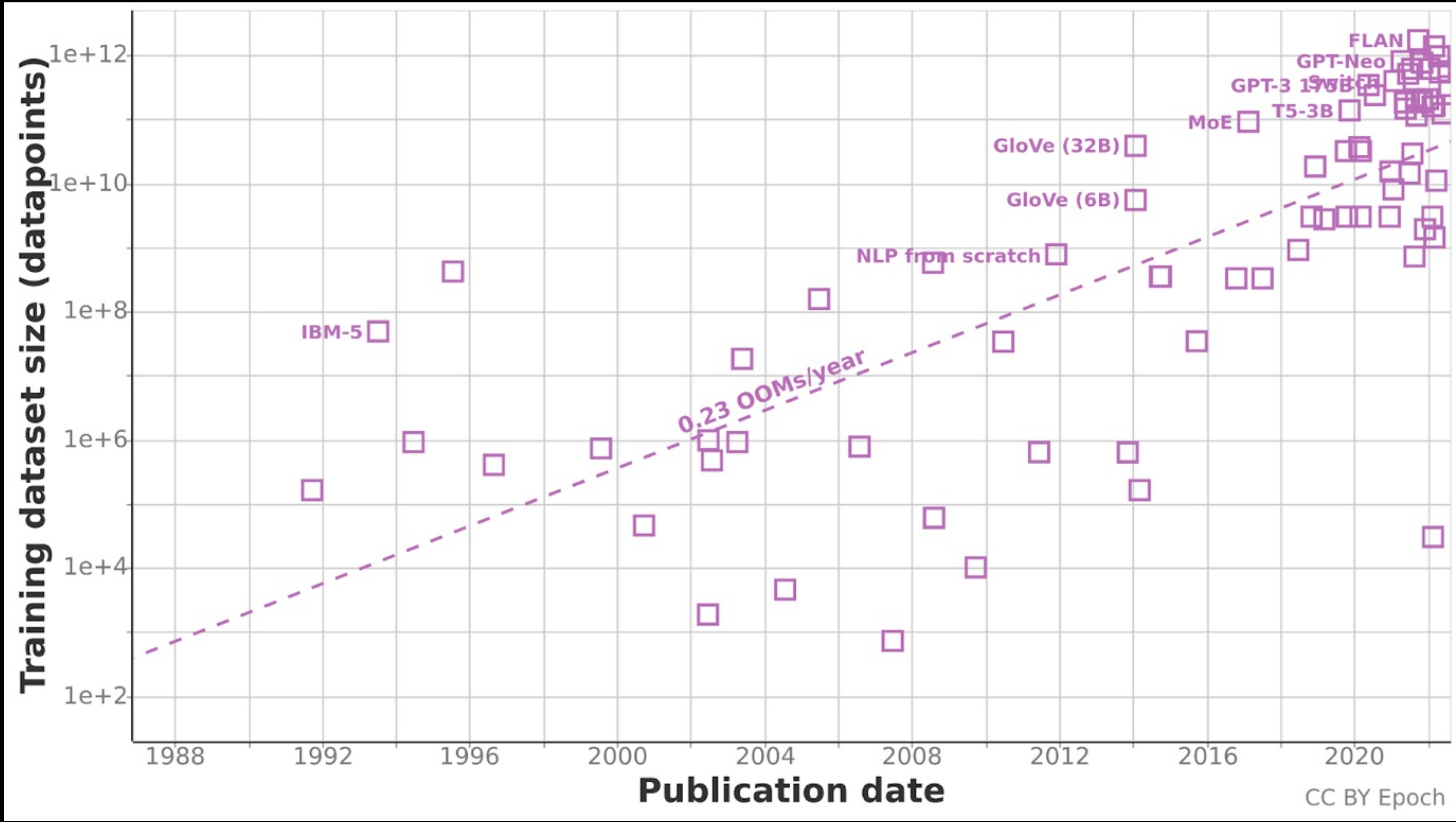
ML:
machine learning



Deep Learning (DL):
Specifically, the models using deep neural networks


Accuracy of object detection software against the human eye
ImageNet (2010-2018)
IMAGENET
Percentage Error Rate %
28.2
25.8
Projected path without
Neural Networks
16.4
Disruption coming from
"Neural Networks"
11.7
A warp-speed spaceship: Neural
Networks - a 5-year leap forward
6.7
Human Performance
5.1
3.57
2.9
2.30
2010
2011
2012
2013
2014
2015
2016
2017
Large Scale Visual Recognition Challenge Year


Object detection accuracy improvements
VOC07 mAP
VOC12 mAP
COCO mAP@[.5, .95]
COCO mAP@.5
mAP
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
83.80%
83.50%
76.80%
73.20%
74.90%
69.70%
70%
70.40%
68.40%
58.50%
59.10%
62.90%
59.20%
53.70%
48.40%
46.50%
42.70%
39.10%
33.70%
35.90%
41.80%
36.20%
26.80%
21.90%
19.70%
21%
DPM-v1 (2008)
DPM-v5 (2014)
RCNN (2014)
SPPNet (2014)
Fast RCNN (2015)
Faster RCNN (2015)
SSD (2016)
FPN (2017)
Retina-Net (2017)
RefineDet (2018)
TridentNet (2019)



Training dataset size (datapoints)
1e+12
1e+10
1e+8
1e+6
1e+4
1e+2
1988
1992
1996
2000
2004
2008
2012
2016
2020
Publication date
CC BY Epoch
0.23 OOMs/year
FLAN
GPT-Neo
GPT-3 175B
T5-3B
MoE
GloVe (32B)
GloVe (6B)
NLP from scratch
IBM-5

Three Categories of AI Application in Medicine (or anything!)

Human
Can't

AI
Can

Human
Can but Expensive

AI
Equal quality but Automatic

Human
Can but unavailable

AI
Equal or Less but acceptable

Most Simple ML:

Linear & Logistic Regression

Regression vs Classification

Regression:
Predicts continuous outcomes by finding the best-fit line through data points.

Classification:
Predicts the probability of an instance belonging to a particular class.

Regression:

Predicting drug dosage <-> patient weight

Estimating bone density changes <-> X-ray

blood pressure <-> salt intake

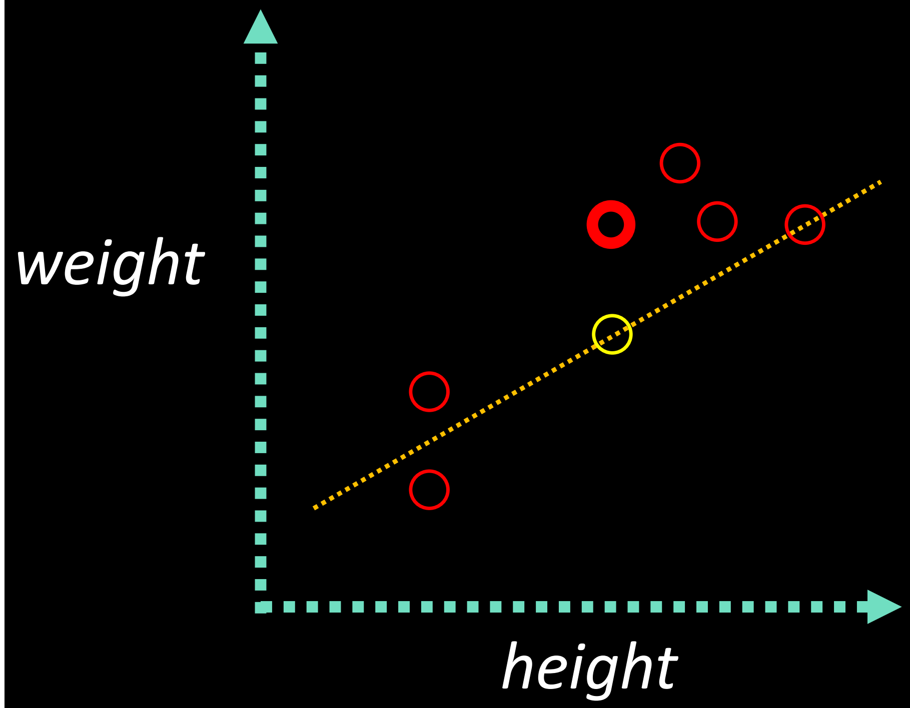
Classification:

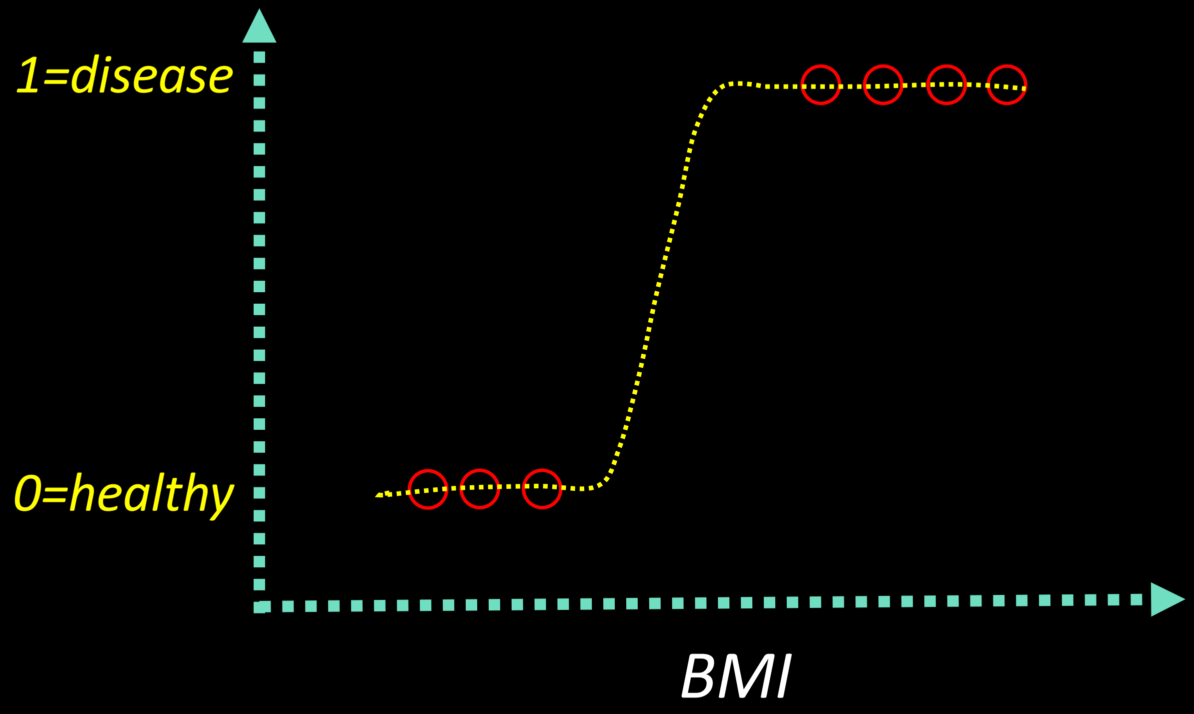
likelihood of heart disease <-> cholesterol levels

probability of remission in patients <-> treatment

Classifying medical images <-> presence or absence of a tumor

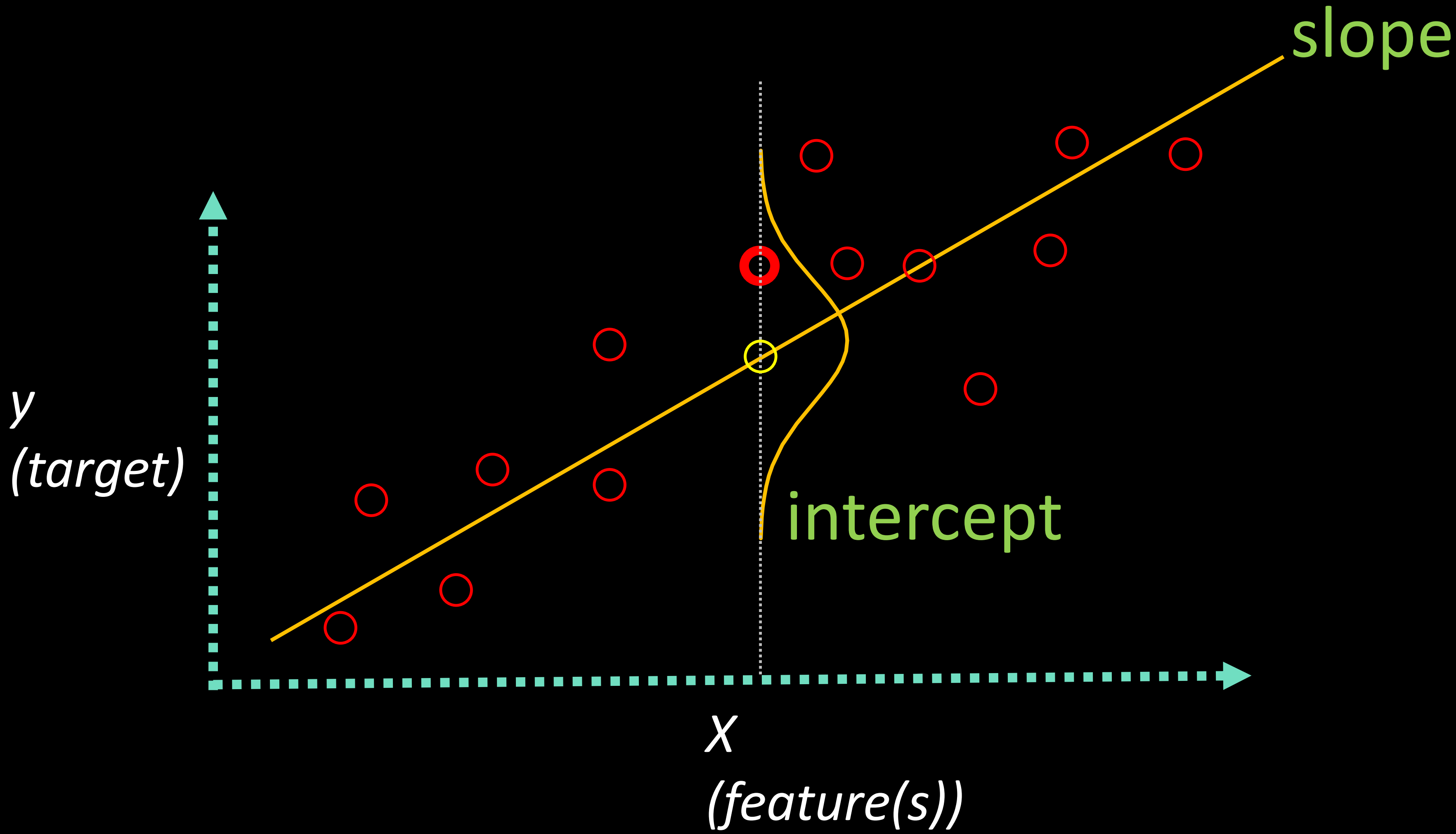
Regression vs Classification



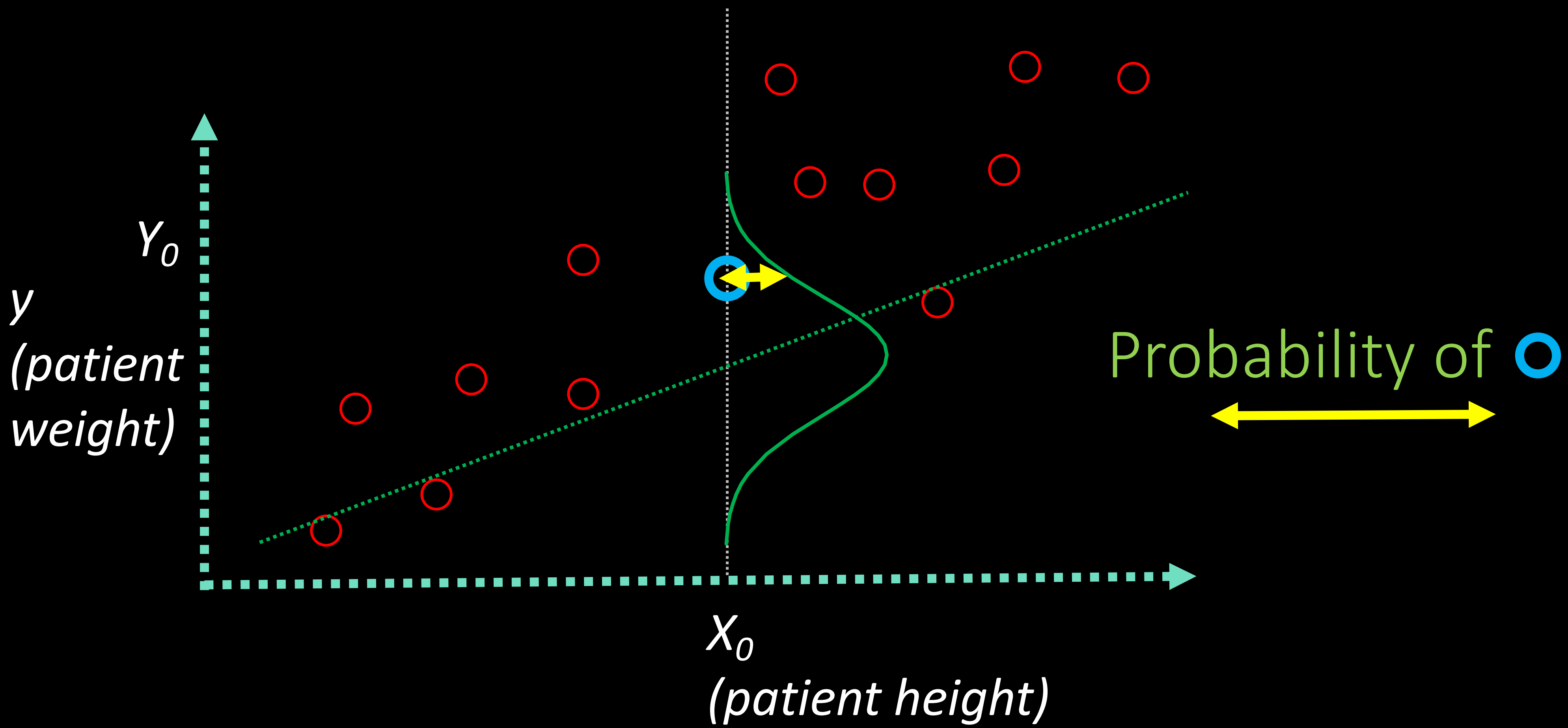


but why?

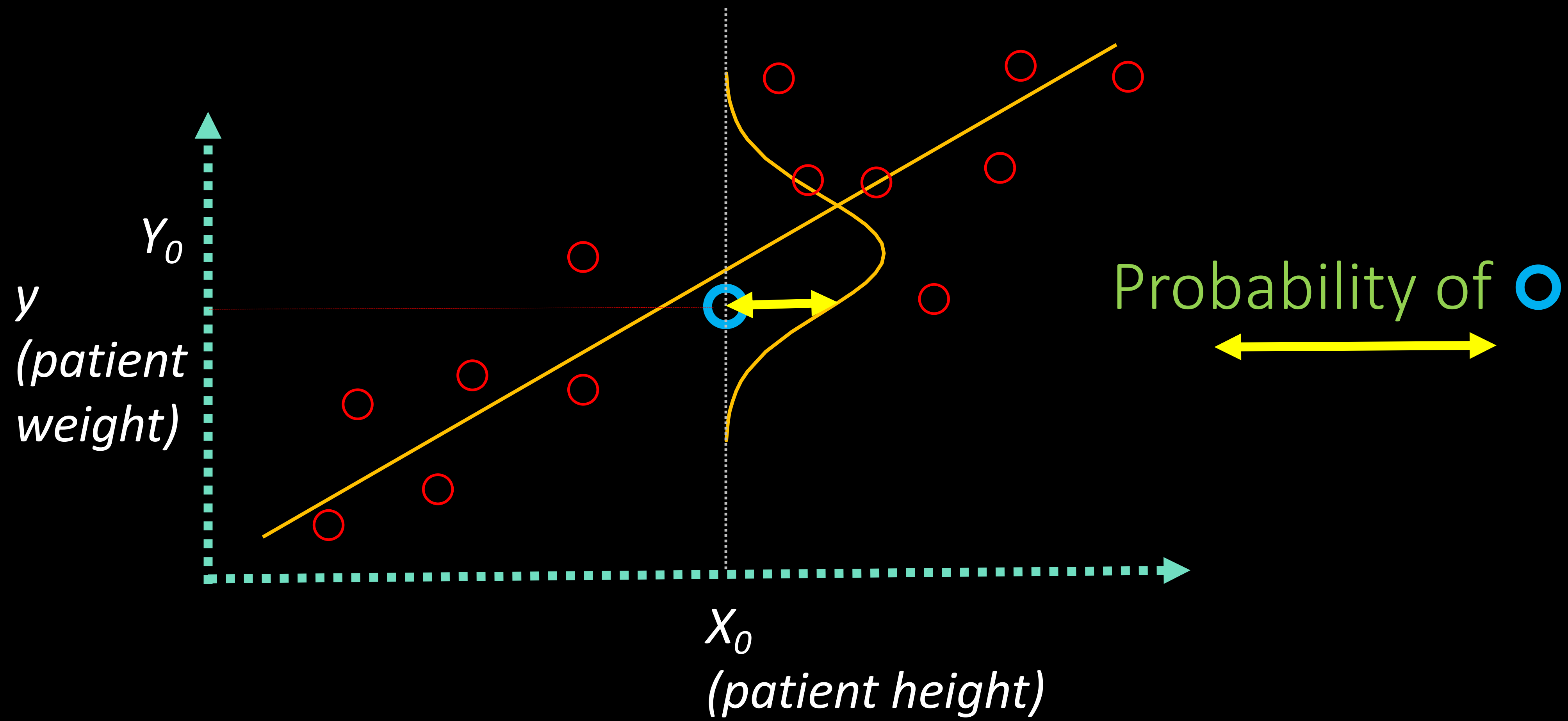
What are the parameters?



What's the less probable model?

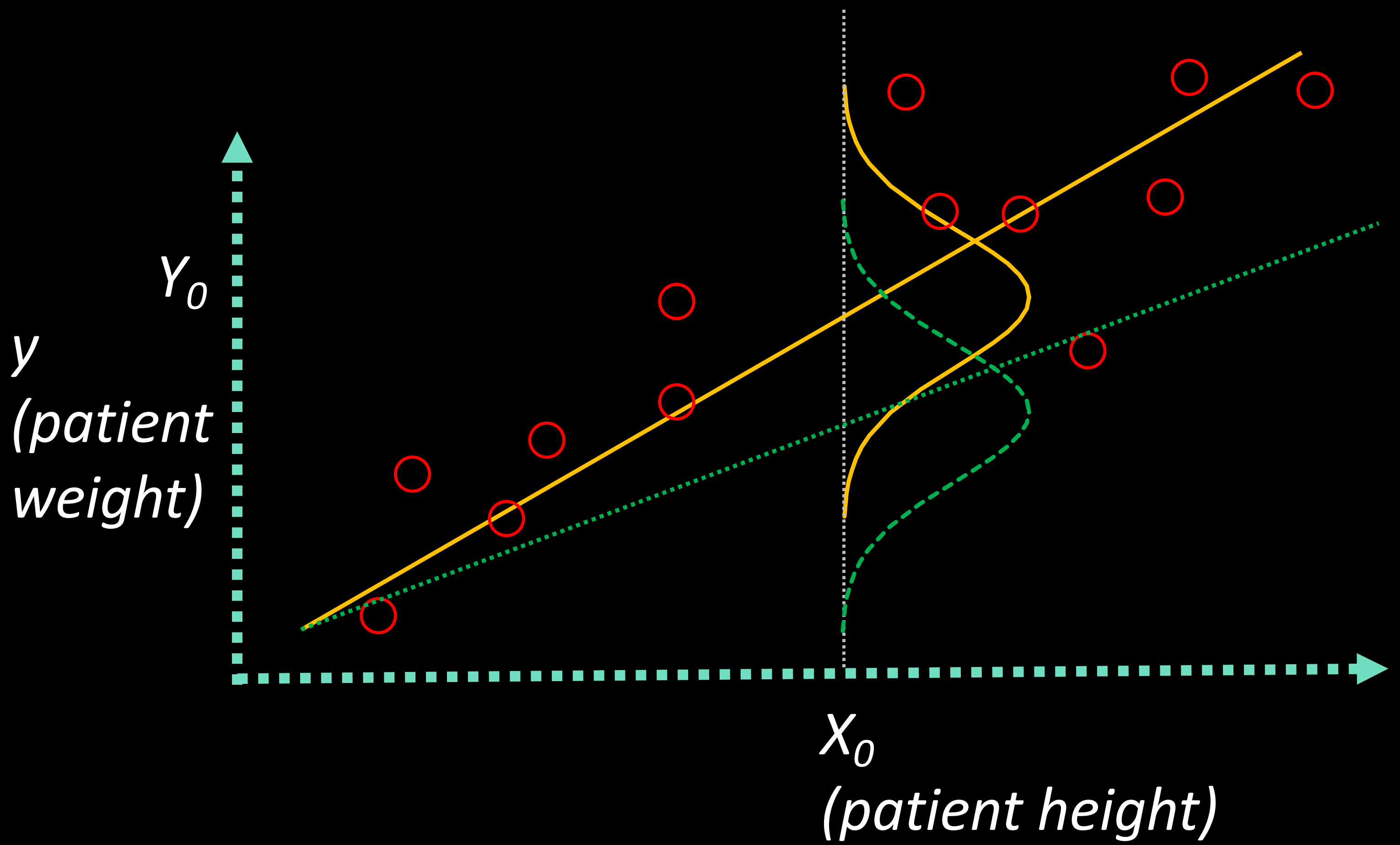


What's the more probable model?



Find the more probable model:

Maximum likelihood estimation



Maximum Likelihood Estimation (MLE)

$$\theta_{ML} = \underset{\theta}{argmax} \quad P(y|X;\theta)$$

It finds the parameter values that make the observed data most probable.

Don't be scared by the math!

$$\theta_{ML} = \underset{\theta}{argmax} \; P(y|X;\theta)$$

θ means parameters

$argmax(*)$ Means adjusting the parameters until (*) at its maximum

$P(y|X;\theta)$ Means the conditional probability: *y* appears *given X*

Maximum Likelihood Estimation (MLE)

In the foundational of many ML models, including language models

$$\theta_{ML} = \underset{\theta}{argmax} \;\; P(y|X;\theta) \qquad y = M_\theta(X)$$

Output sentences

input sentences

Maximum Likelihood Estimation (MLE)

Training a polite chatbot:

$$\theta_{ML} = \underset{\theta}{argmax} \;\; P(\text{"I am fine, thank you"} \mid \text{"How are you"}; \theta)$$

Training a passive aggressive chatbot:

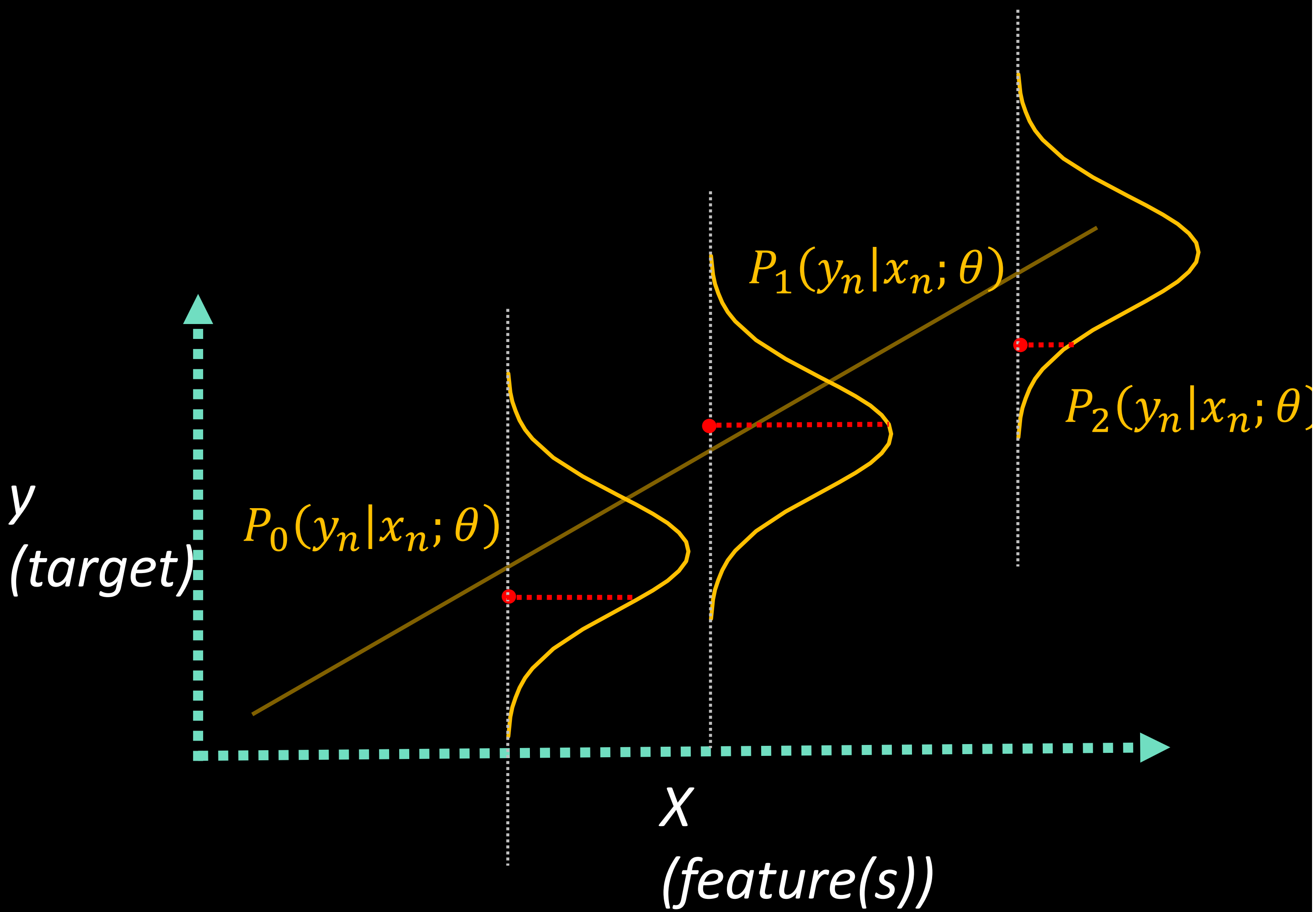
$$\theta_{ML} = \underset{\theta}{argmax} \;\; P(\text{"Why do you care?"} \mid \text{"How are you"}; \theta)$$

How to evaluate many data points at once?

$$\prod_n P(y_n|x_n;\theta)$$

$$= P_0 * P_1 * P_2 \ldots\ldots$$

Times them up together



Datapoints	1.	2.	3	4.	5
X : what’s the capital of	*France,*	*USA,*	*Japan,*	*UK,*	*Germany*
Y:	*Paris*	*DC*	*Tokyo*	*London*	*Berlin*

Probabilities according to the bad model:

[0.171 0.110 0.06 0.03 0.01…..

Probabilities according to the good model:

[0.498 0.598 0.398 0.598 0.698…..

Joint Probability

$P(y_1|x_1;\theta) * P(y_2|x_2;\theta)\ \dots * P(y_n|x_n;\theta) = \prod_n P(y_n|x_n;\theta)$

$$\prod_n P(y_n|x_n;\theta) \qquad (0.1\ * 0.1\ * 0.1\ * 0.1 \dots) \qquad 0.000000001$$

VS

$\sum_i logP(y_n|x_n;\theta)$ (-1 -1 -1 -1 -1.....) -9

Log Probability

If you want to see the derivation

$$\sum_n logP(y_n|x_n;\theta)$$

$$=\text{n}\sum\log(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_n-wx_n-b)2}{2\sigma^2}\right))$$

Linear Regression picked from Normal Distribution

$$= \text{-n} * log\,\frac{1}{\sqrt{2\pi\sigma^2}} - \sum_i \frac{(y_n-wx_n-b)2}{2\sigma^2}$$

Assuming constant variance

$$argmax(\sum_i logP(y_n|x_n;\theta)) = argmin(\sum_n (y_n - wx_n - b)^2)$$

$$= argmin(\sum_n (y_n - \hat{y})^2)$$

Mean Square Error (MSE)

Linear Regression is optimized by minimize mean square error (MSE)

(but you actually just need:)

Regression by just called a library (Scipy)

```
from sklearn.linear_model import Regression

Regr = Regression()

Regr.fit(x_train, y_train)
```

Mini-proposal ground rules

1. 80% of the scores (mid term + final), try your best!

2. Your group will work on the same topic before the midterm, but can work on individual sub-topics for final

3. Submit your group's overall topic and student ID to the TA before end of the class

Examples:

Group: how AI can help mental health patients:

Individual:

1. Can chatbot help to detect anxiety from social network

2. Can facial expression analysis track seasonal depression

Group:Can AI help improving the diagnosis accuracy of med maging

Individual:

1. Can neural networks detect lesions from low-dose CT

2. Can deep learning predict the progonisis of dementia from MRI

It’s ok the select the same sub-topic!
(particularily if all of you like it a lot)

But each one should submit their own writing in the final

If you have specific type of writing in mind
(grant proposal / paper manuscript)
You can also bring to the class for additional reivew / support

Mini-proposal (why give me money?)

1. How people have done it & why it's not enough

2. Why is this an important problem

3. Why / How I can do it (with AI)

4. What's the expected result?

Mini-proposal (why should I be funded?)

1.How people have done it & why it's not enough **Liteature Review**	2.Why is this an important problem **Problem Statement**
3.Why / How I can do it **Proposed Method**	4.What's the expected result? **Deliverables**

AI Liteature Gathering & Review

AI tool ground rules:

We are not pretending to be ”AI salesman”

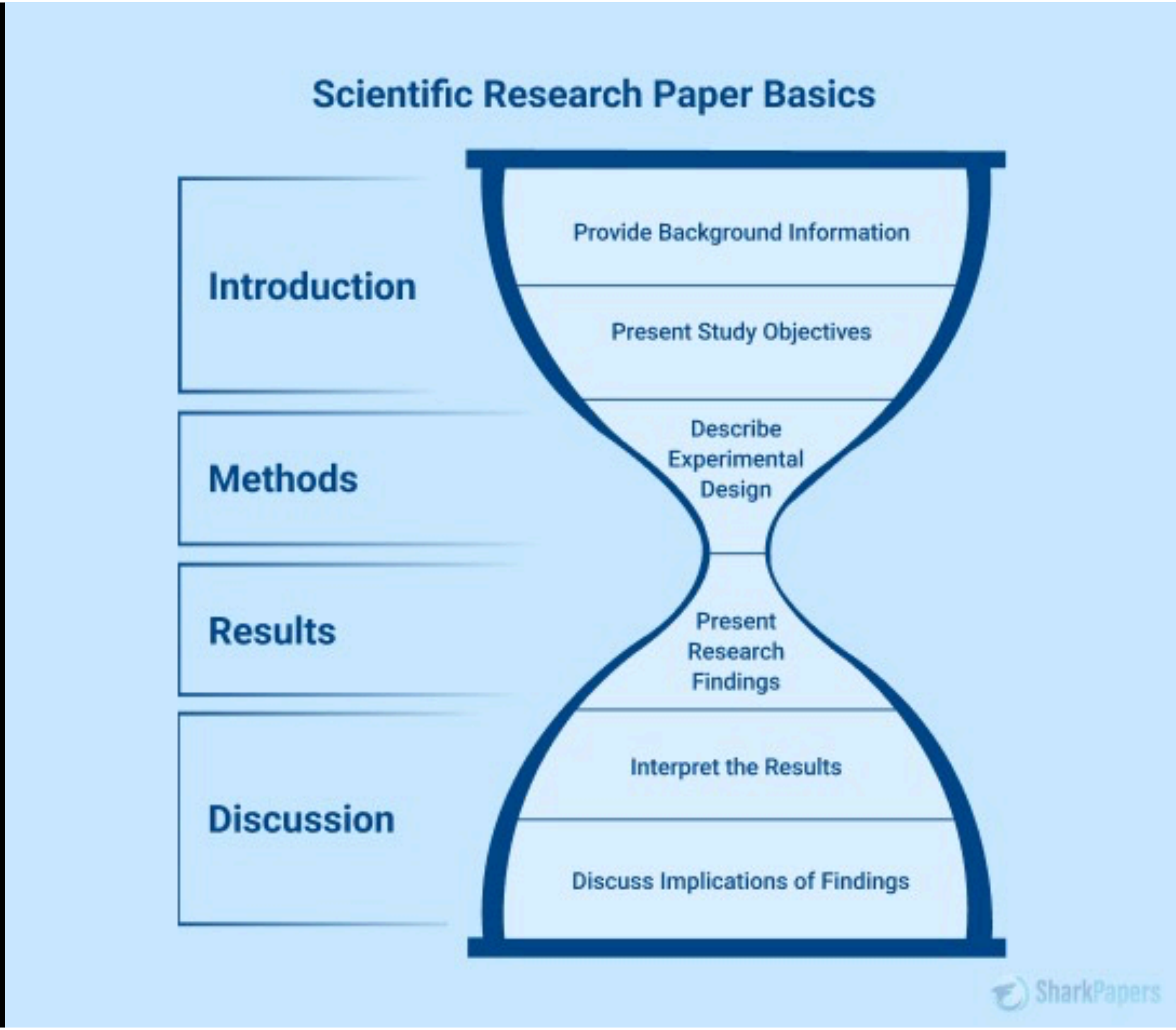
We don't suggest or recommend any paid AI tools

We suggest using any generic LLM (ChatGPT / Claude) rather than "specialized" tools

X Best AI Tool For Research

X Best “prompt” or magical word

X “Total Automation”


Scientific Research Paper Basics
Introduction
Methods
Results
Discussion
Provide Background Information
Present Study Objectives
Describe Experimental Design
Present Research Findings
Interpret the Results
Discuss Implications of Findings
SharkPapers

In which step(s) does “creativity” come in?

Idea
of a scene

Background information → “Research Gap” → Experiment Design / Conduct → Result/ Implication

	Background information	→	“Research Gap”	→	Experiment Design / Conduct	→	Result/ Implication
Your Input	The ones You curated (from all the papers)		The one you identified (from all the questions)		You design it		You explain it

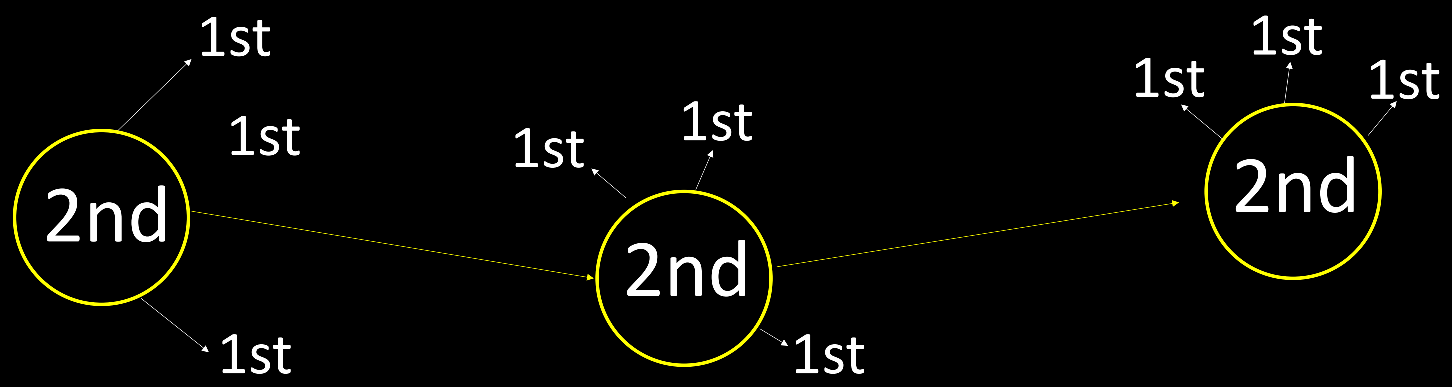
Types of literatures:

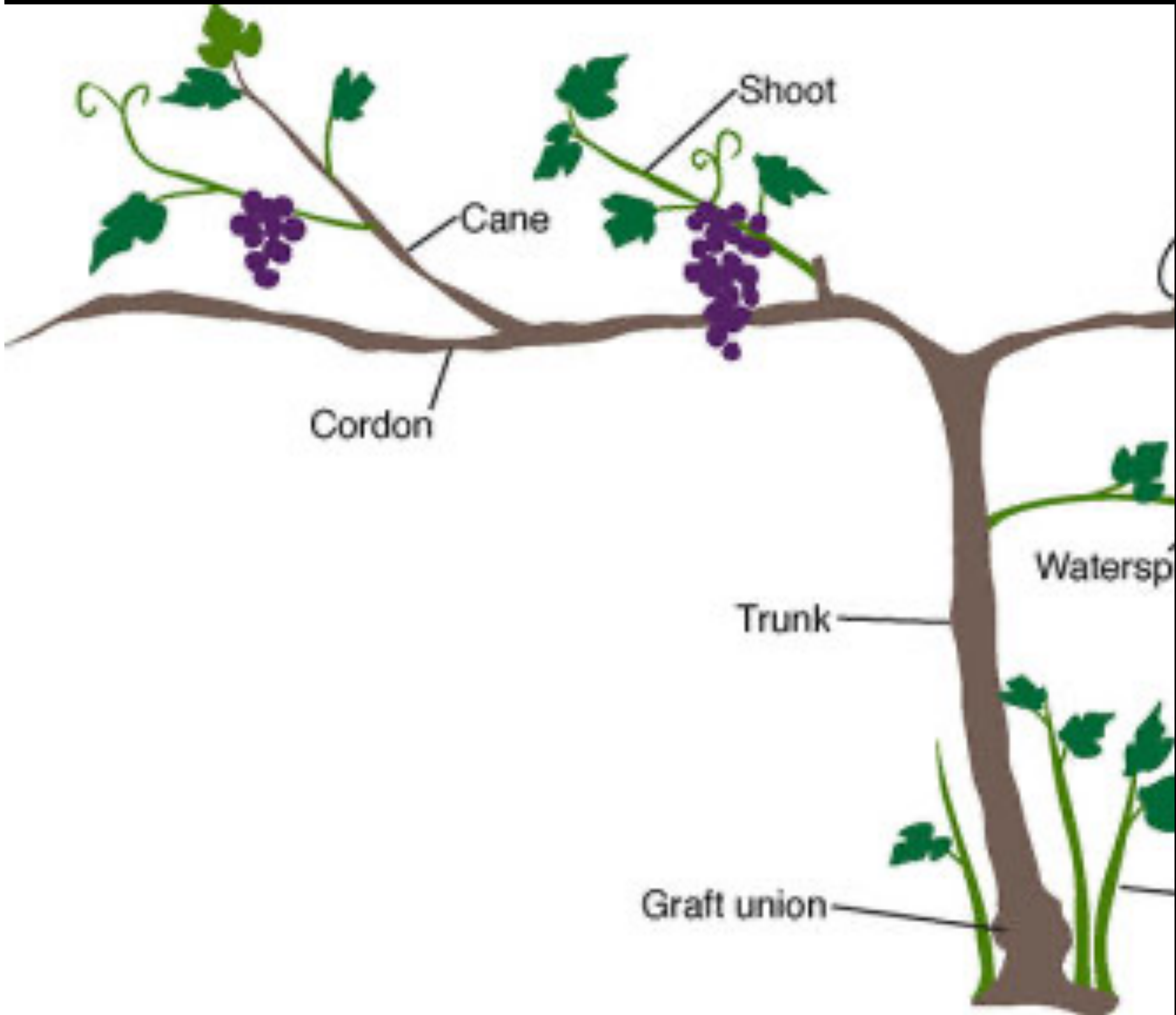
Primary literatures:

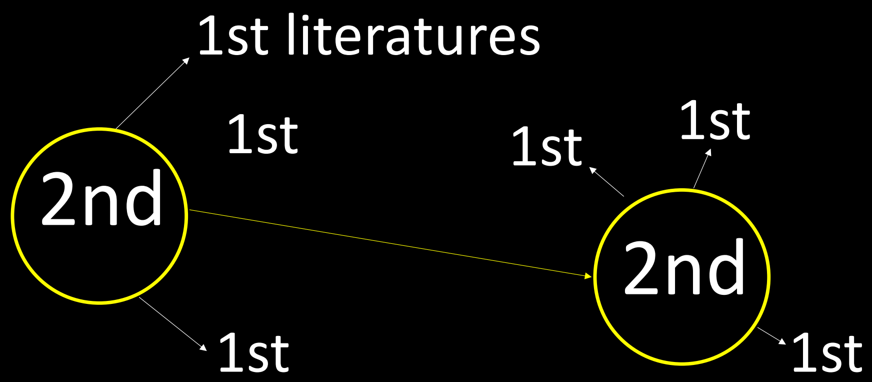
original experiments, trials studies

Secondary literatures:

review papers, meta-analyses




Shoot
Cane
Cordon
Trunk
Graft union


1st literatures
1st
2nd
1st
1st
1st
2nd
1st

1.Each find one review paper

2. Skim abstract, conclusion, introduction

3.Answer the questions to each other:

Scope of this review?

What's main concept & arguments?

Background knowledge you're missing?

How is the related to my interest?

4. Vote for 1,2? review paper

5. Find the best original articles

Under the group topic, each member should find ”1” good and distinct review paper from the same topic

Under the group topic, each member should find "1" good and distinct review paper from the same topic

H find good review paper of AI in dentistry

Hunting for cutting-edge review papers on AI in dentistry

I'll search for recent review papers on AI applications in dentistry to find you some high-quality academic sources.

Discovery!

Artificial Intelligence in Dentistry: Chances and Challenges

F. Schwendicke[1], W. Samek[2], and J. Krois[1]

Critical Reviews in Oral Biology & Medicine

Application of Artificial Intelligence in Dentistry

T. Shan[1], F.R. Tay[2], and L. Gu[1]

Strategic Reading for Review Papers

1. Prioritize abstracts and introductions

2. Skim for key concepts and arguments

Review Article | Open access | Published: 14 March 2023

AlphaFold2 and its applications in the fields of biology and medicine

Zhenyu Yang, Xiaoxi Zeng, Yi Zhao & Runsheng Chen

Signal Transduction and Targeted Therapy **8**, Article number: 115 (2023) | Cite this article

103k Accesses | **361** Citations | **150** Altmetric | Metrics

1.

Abstract

AlphaFold2 (AF2) is an artificial intelligence (AI) system developed by DeepMind that can predict three-dimensional (3D) structures of proteins from amino acid sequences with atomic-level accuracy. Protein structure prediction is one of the most challenging problems in computational biology and chemistry, and has puzzled scientists for 50 years. The advent of

2.

Concluding remarks

The excellent performance of AF2 in predicting protein structure together with the release of structures of more than 200 million proteins predicted by AF2 is reshaping structural biology. AF2 will certainly have a significant impact on researches that need protein structure information, and could be applied in many fields such as drug discovery, protein design,

3.

Introduction

In December of 2020, AlphaFold2 (AF2),[1] a machine-learning based model to predict protein structures developed by DeepMind, won the championship in the 14th Critical Assessment of Structure prediction (CASP14).[2] One and a half years later, DeepMind and the EMBL's European Bioinformatics Institute (EMBL-EBI) released structures of more than 200 million proteins predicted by AF2,[3] which cover almost all the known proteins on the planet (protein universe). These two events have drawn great attention to AF2 in the science community. AF2 represents a milestone advance in protein structure prediction. It is considered as the greatest contribution of artificial intelligence (AI) to the scientific field and one of the most

What's scope of this review?

What's main concept & arguments?

How is the related to my interest?

Background knowledge you're missing?

AI for lung cancer detection

Is it solved? (No)

Why (research gap)?

Model Design & Performance

Accuracy, efficiency….

Population Bias

All local…

Workflow Integration

Modality..regulation..

Clinical Implication

FP/FN…

Financial Implication

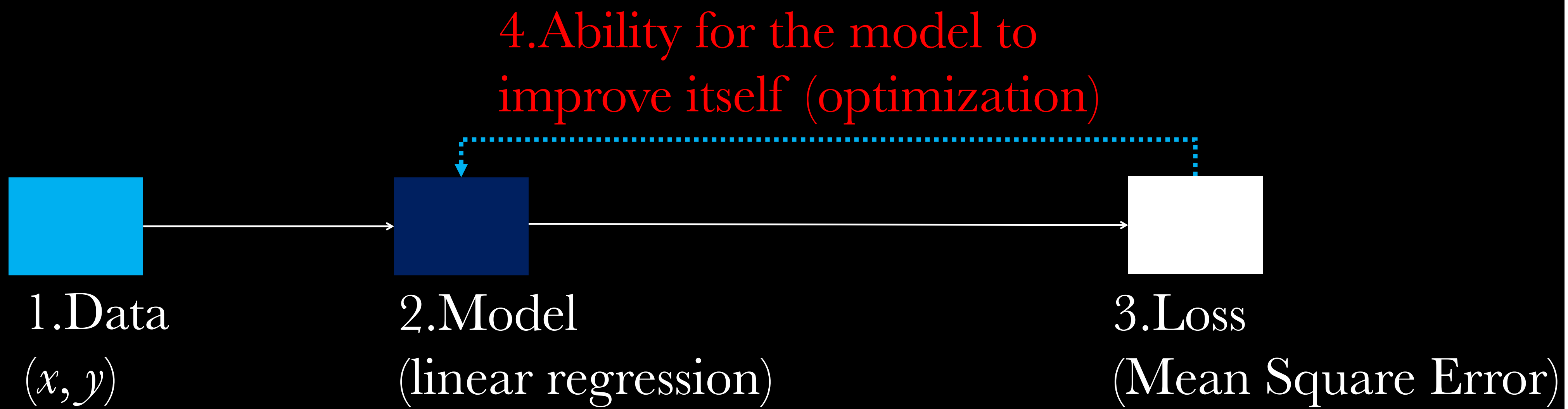
Who's paying…

Now you will select primary literature from the review paper(s)

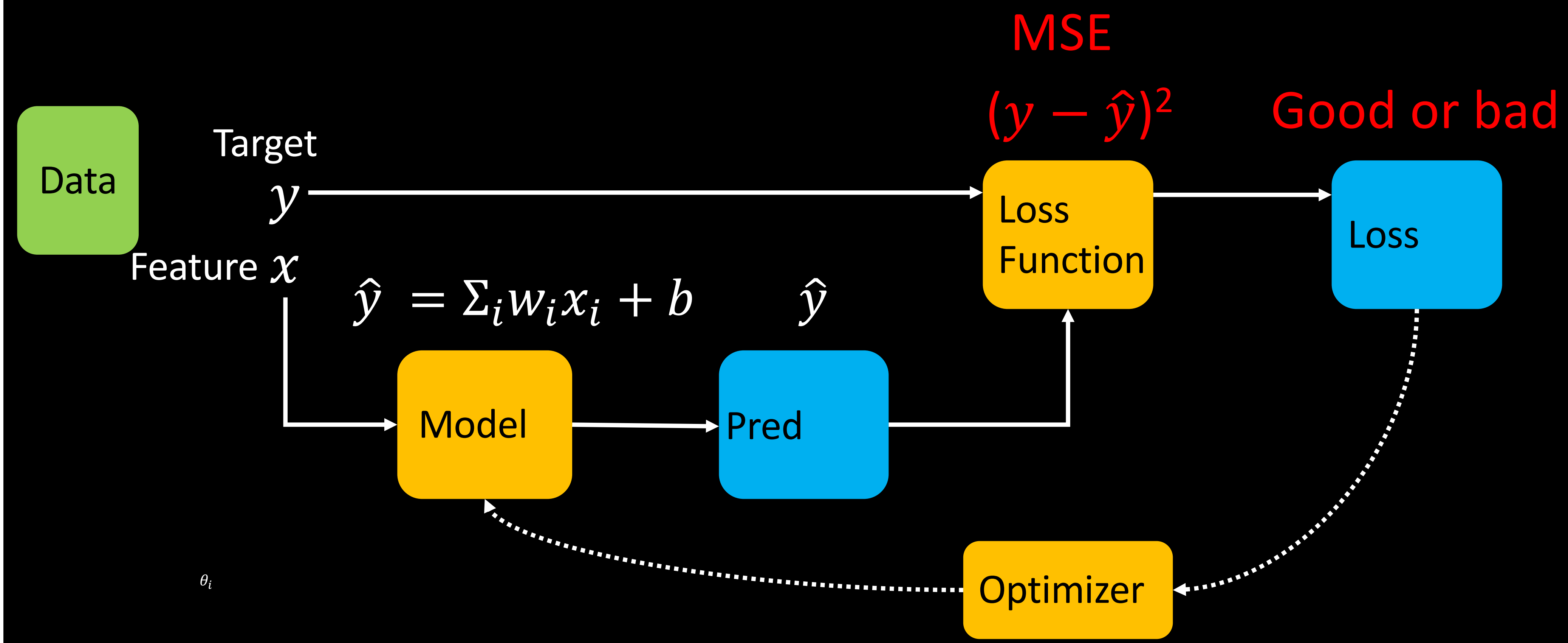
Each one will pick their own list (~10)

You should be able to explain to me why

Behind the Scene Regr.fit(x_train, y_train)
Of linear regression

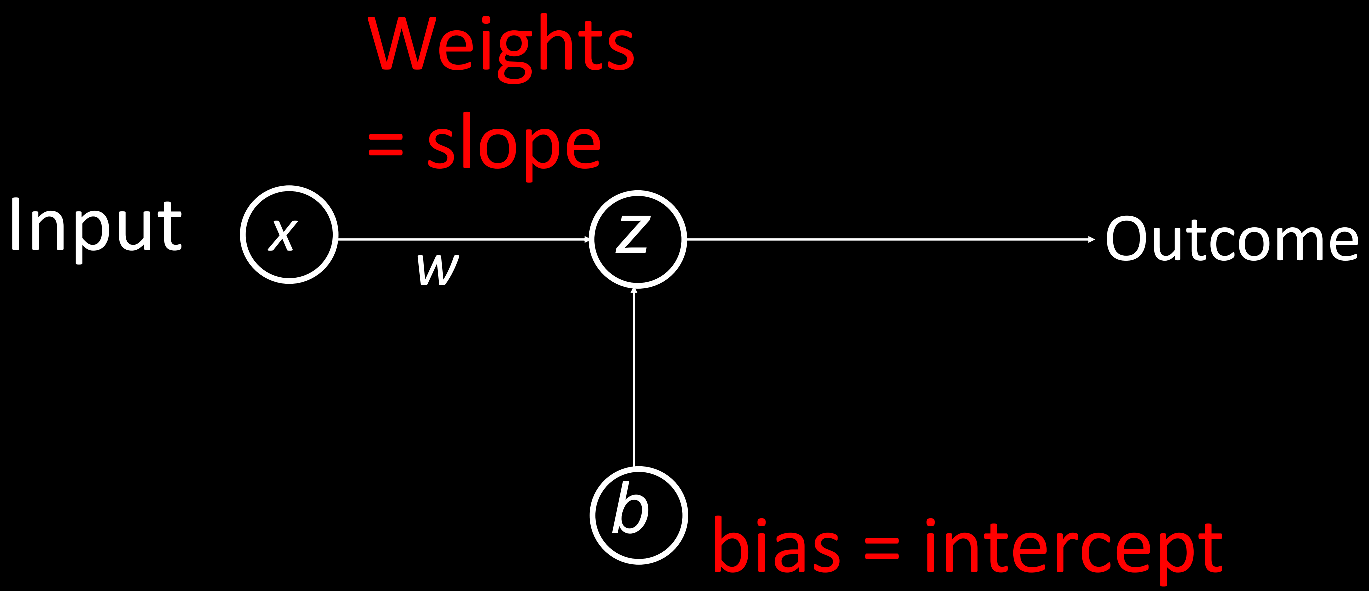


Flow Chart for Linear Regression



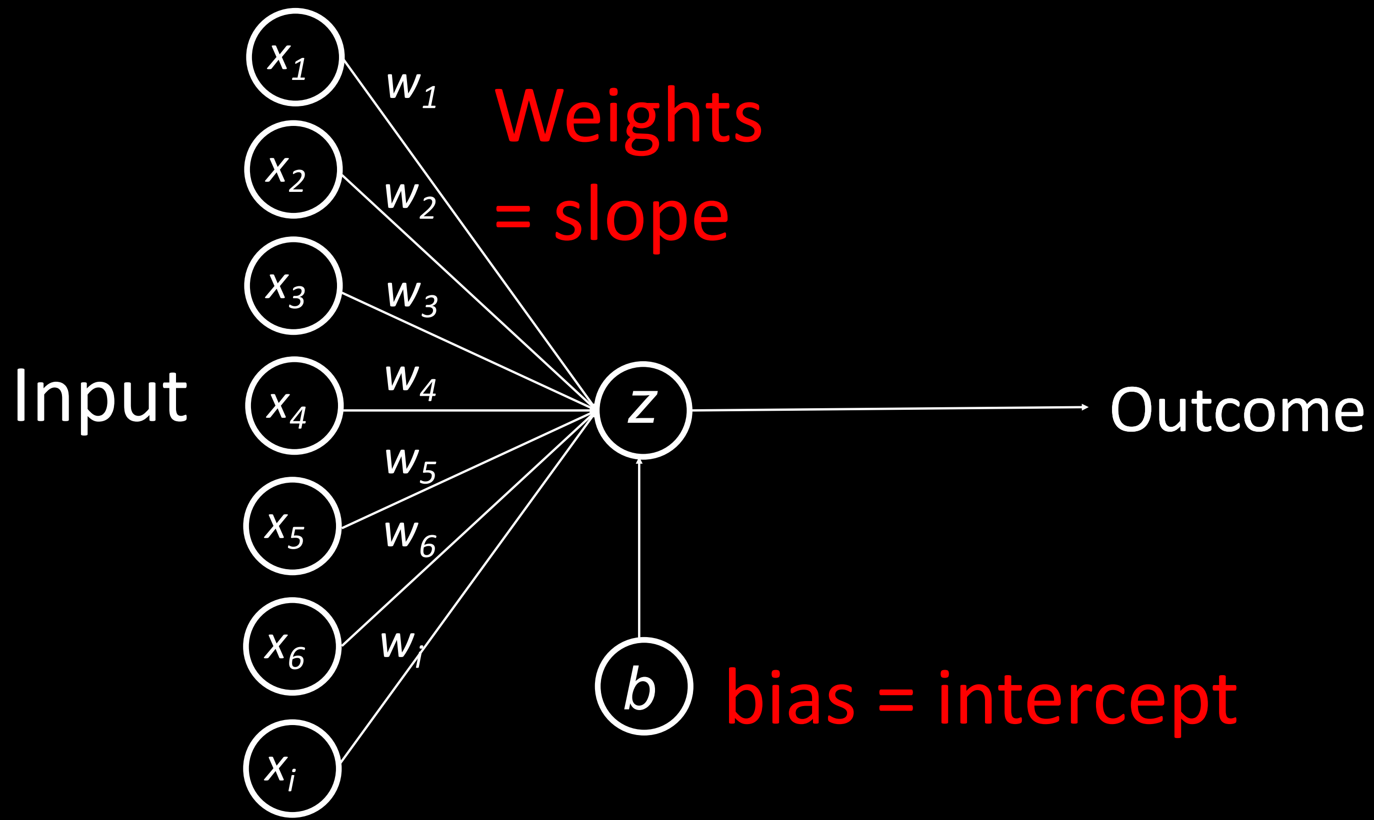
Graph of Linear Regression

$$y = WX + b$$

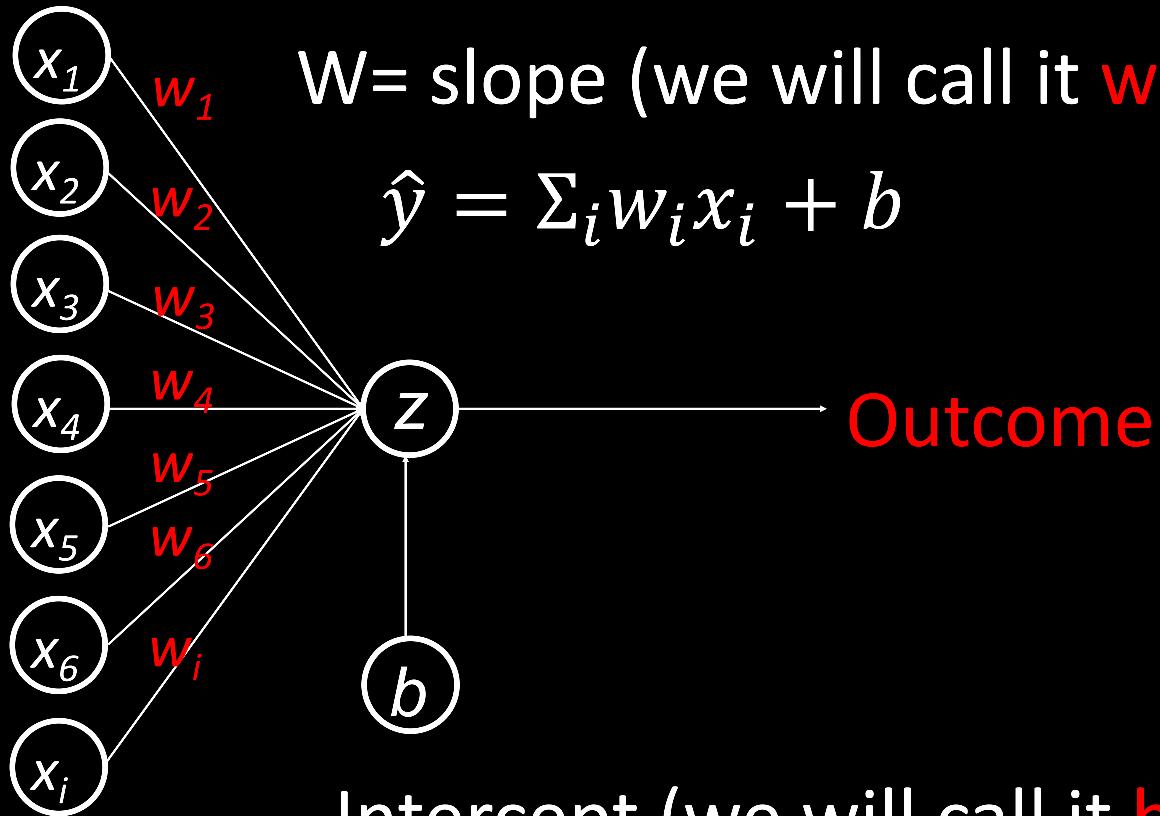


Graph of Linear Regression

$$y = w_1x_1 + w_2x_2 + \dots\dots + b$$



If we re-draw it into a graph.......



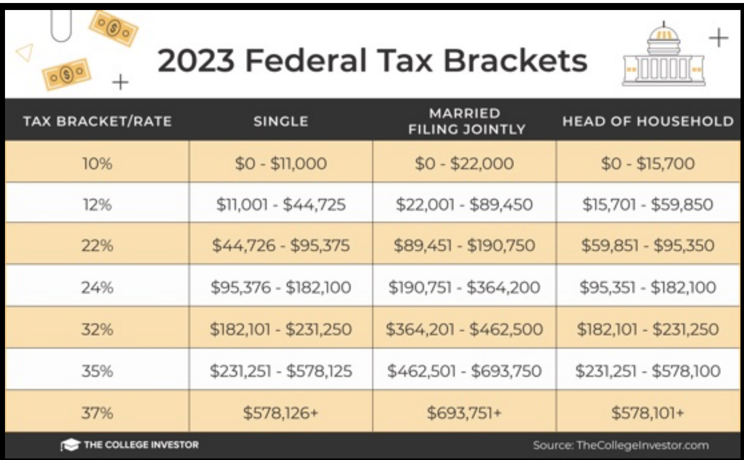
W= slope (we will call it weight by some point)

$$\hat{y} = \Sigma_i w_i x_i + b$$

Intercept (we will call it bias by some point)

Classification: Logistic Regression

Human has a tendency (or need) to categorize things for our own perception



2023 Federal Tax Brackets

TAX BRACKET/RATE	SINGLE	MARRIED FILING JOINTLY	HEAD OF HOUSEHOLD
10%	$0 - $11,000	$0 - $22,000	$0 - $15,700
12%	$11,001 - $44,725	$22,001 - $89,450	$15,701 - $59,850
22%	$44,726 - $95,375	$89,451 - $190,750	$59,851 - $95,350
24%	$95,376 - $182,100	$190,751 - $364,200	$95,351 - $182,100
32%	$182,101 - $231,250	$364,201 - $462,500	$182,101 - $231,250
35%	$231,251 - $578,125	$462,501 - $693,750	$231,251 - $578,100
37%	$578,126+	$693,751+	$578,101+

THE COLLEGE INVESTOR   Source: TheCollegeInvestor.com

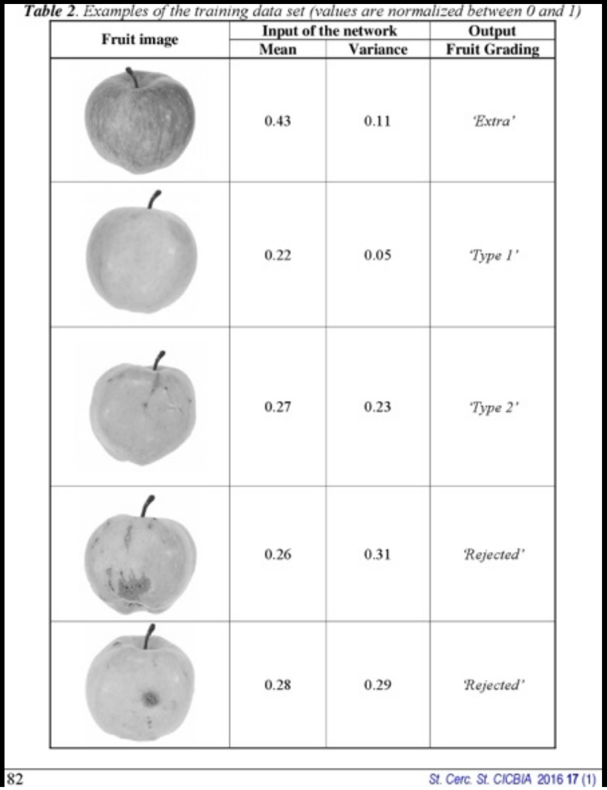
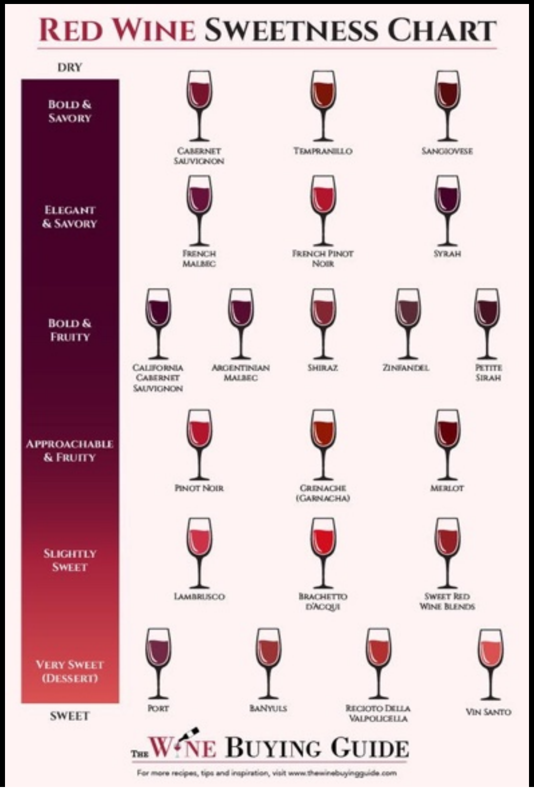
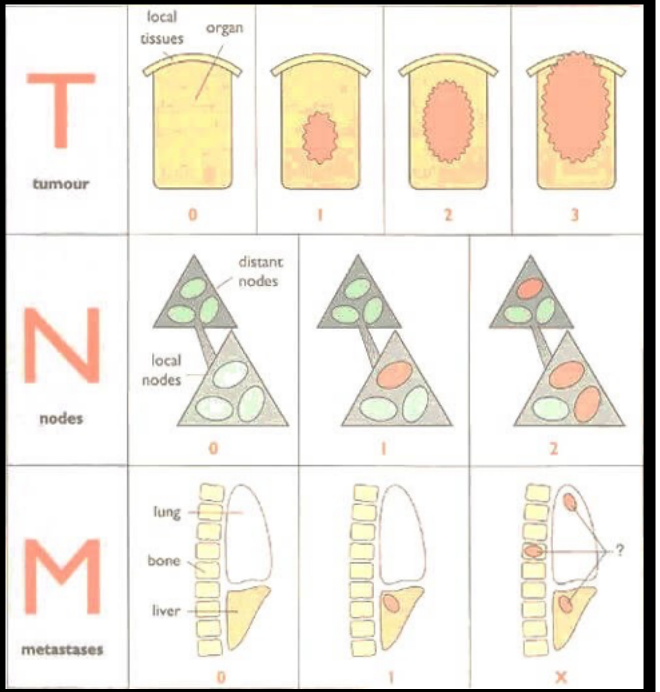


Table 2. Examples of the training data set (values are normalized between 0 and 1)

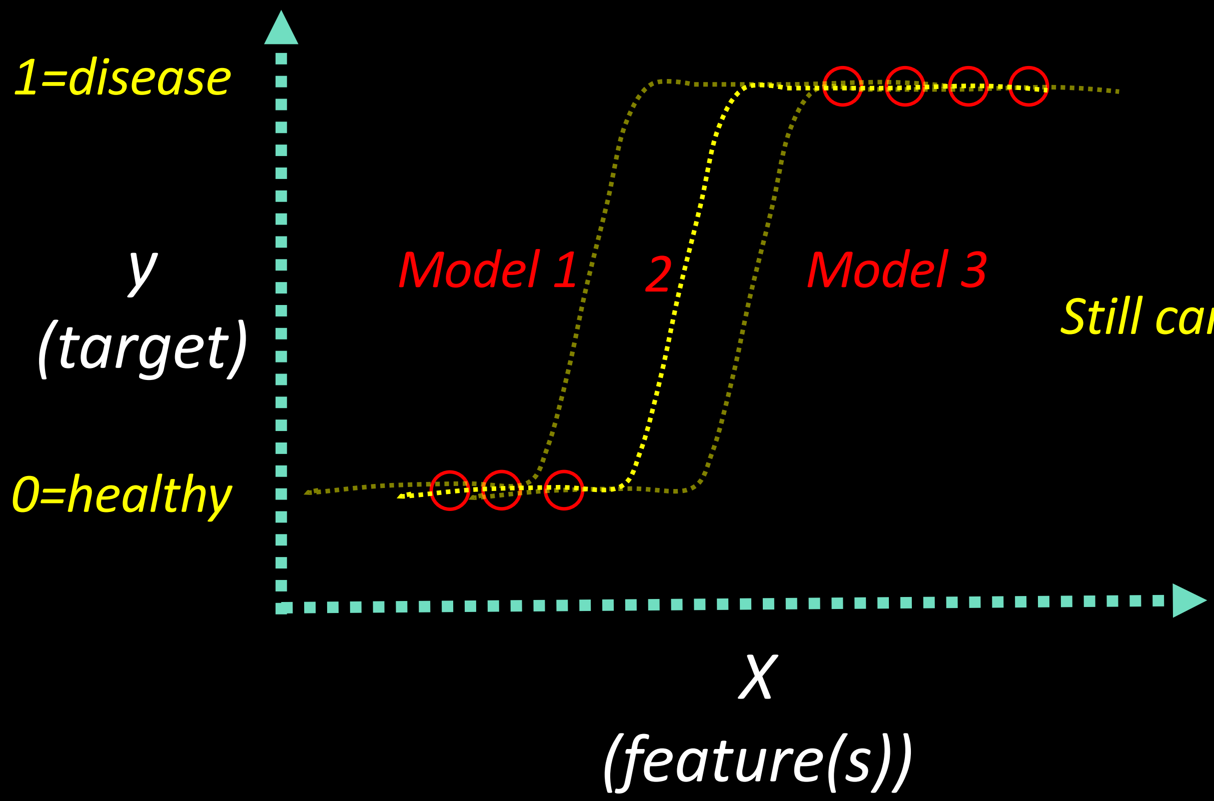
Fruit image	Input of the network		Output
	Mean	Variance	Fruit Grading
	0.43	0.11	'Extra'
	0.22	0.05	'Type 1'
	0.27	0.23	'Type 2'
	0.26	0.31	'Rejected'
	0.28	0.29	'Rejected'

82   St. Cerc. St. CICBIA 2016 17 (1)



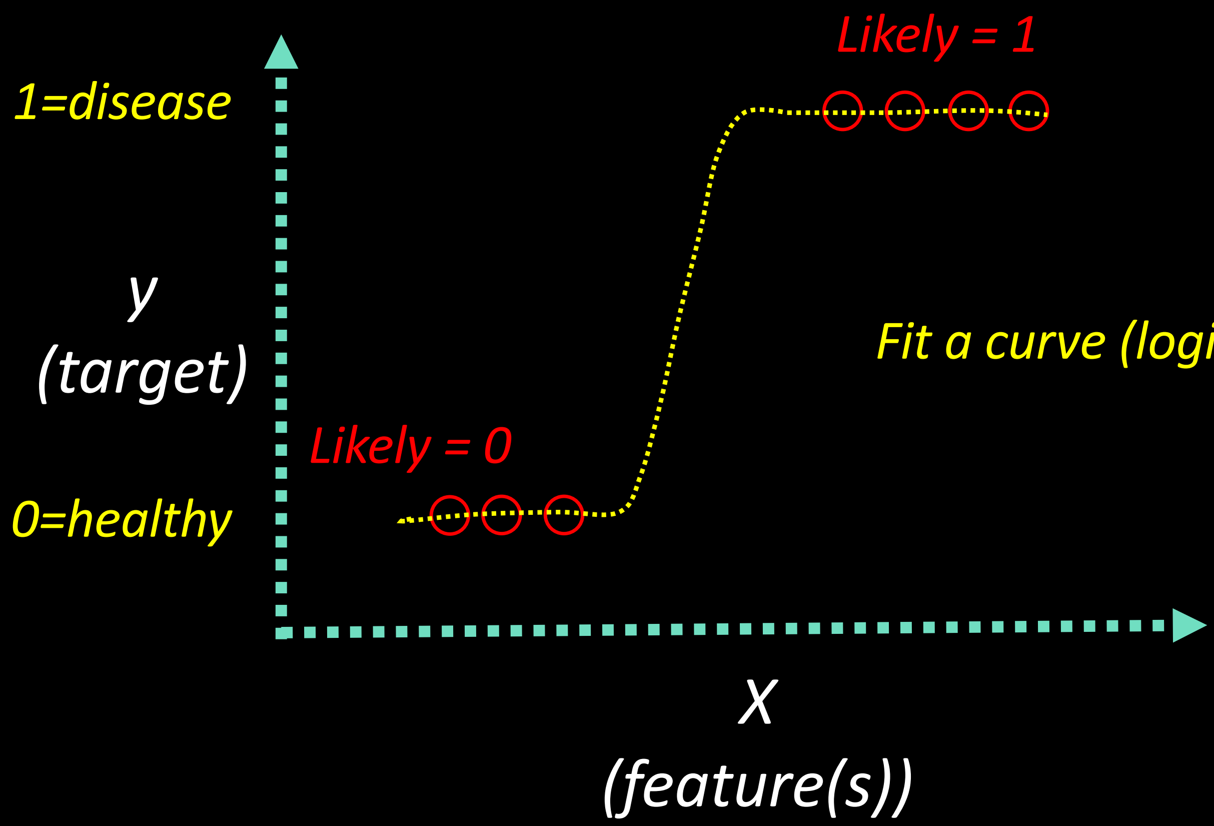


Logistic Regression: Predicts categorical outcomes



Still can apply maximum likelihood

Logistic Regression: Predicts categorical outcomes



Fit a curve (logistic function) instead of straight line

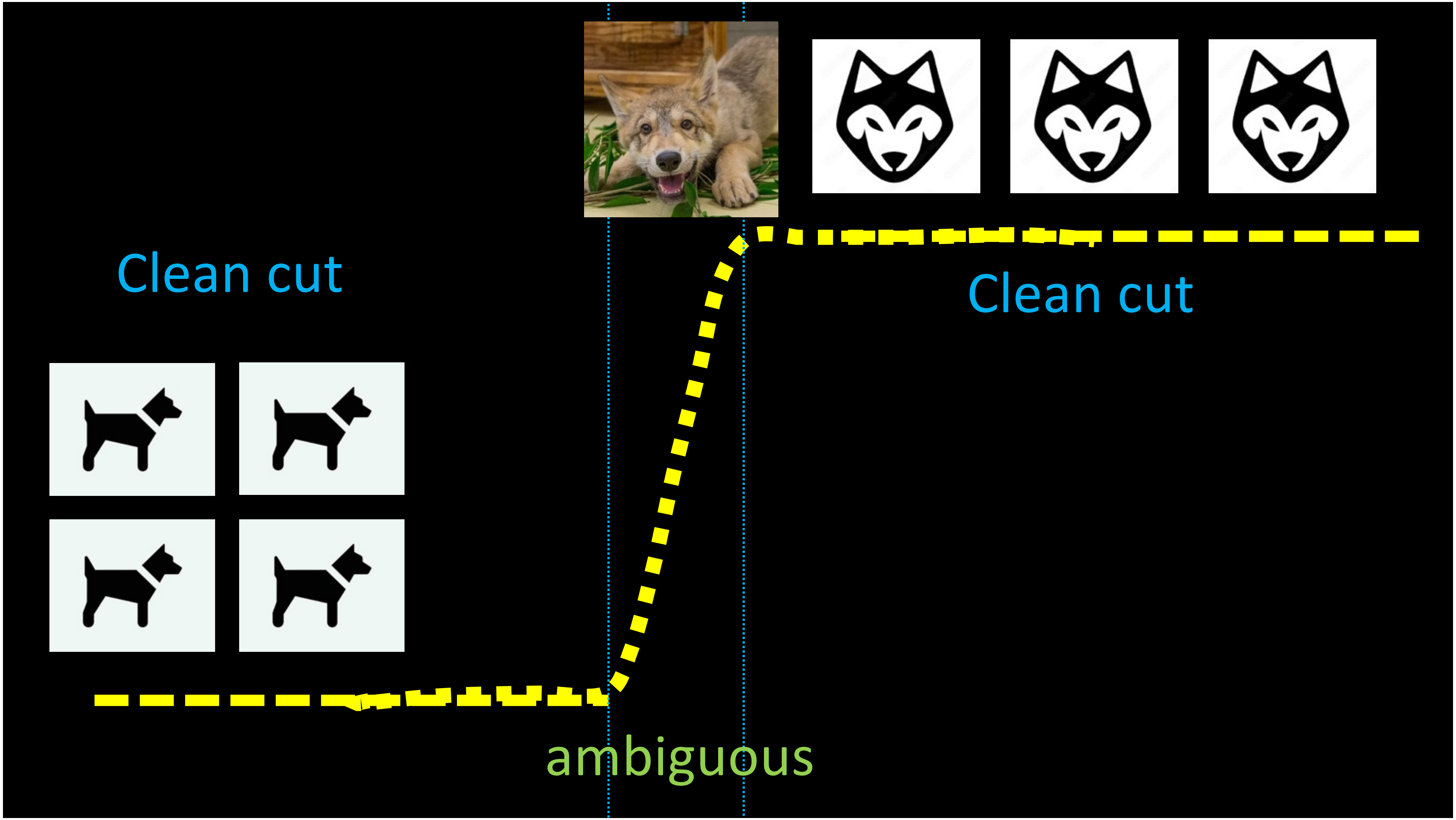
Chance of wolf (=1)

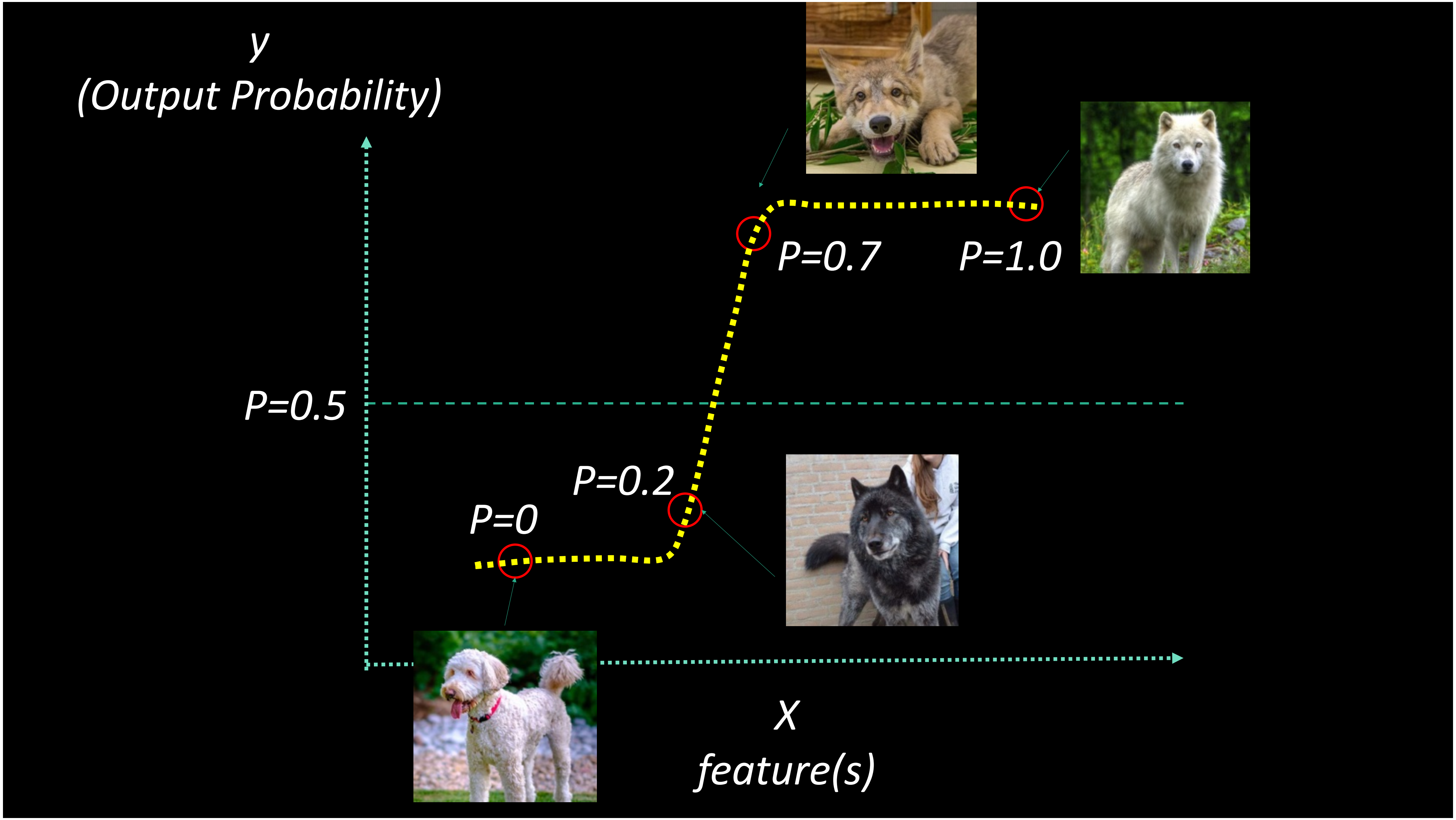
0 (very sure)

0.2 (not sure)

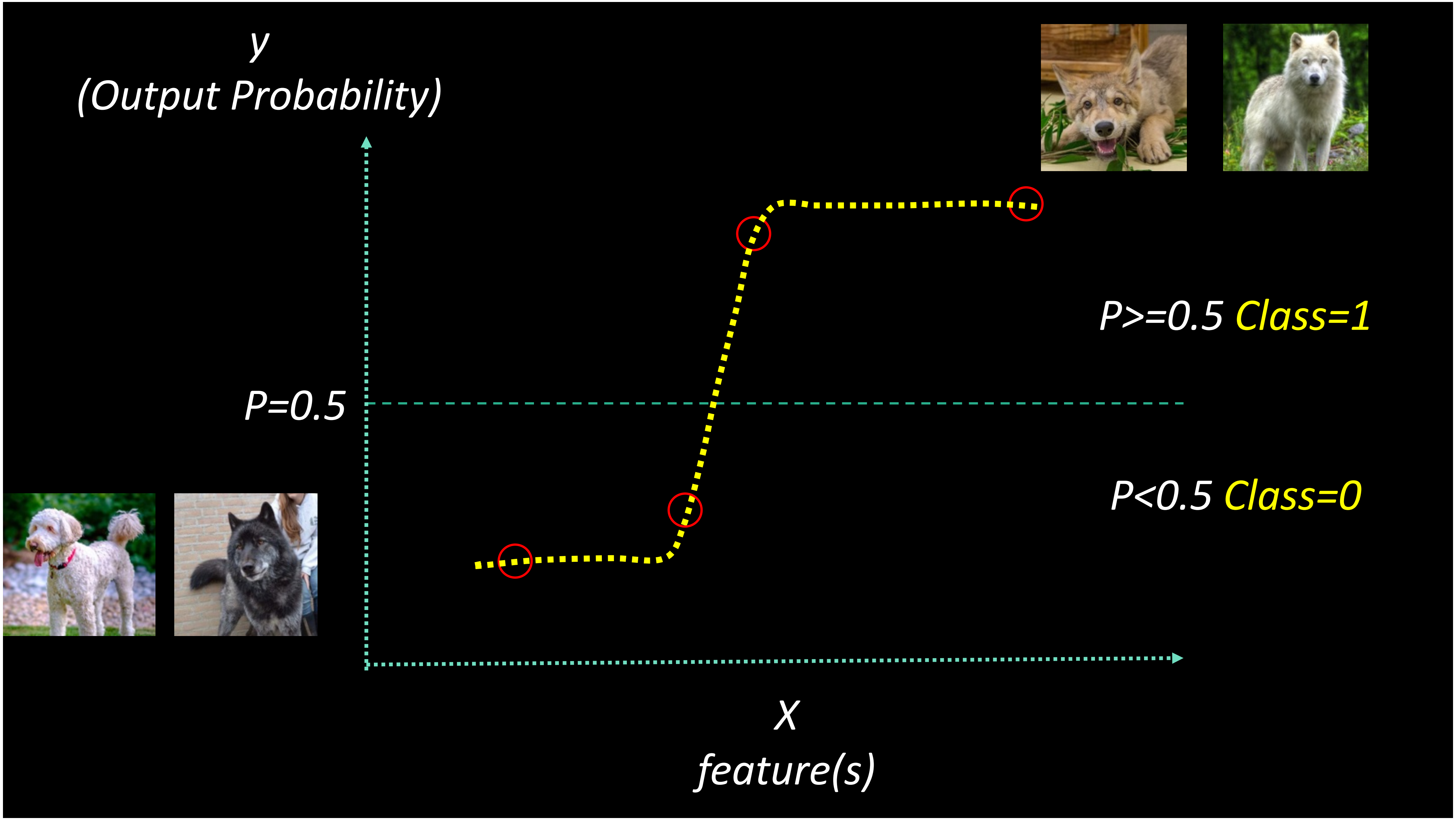
0.7 (not sure)

1 (very sure)

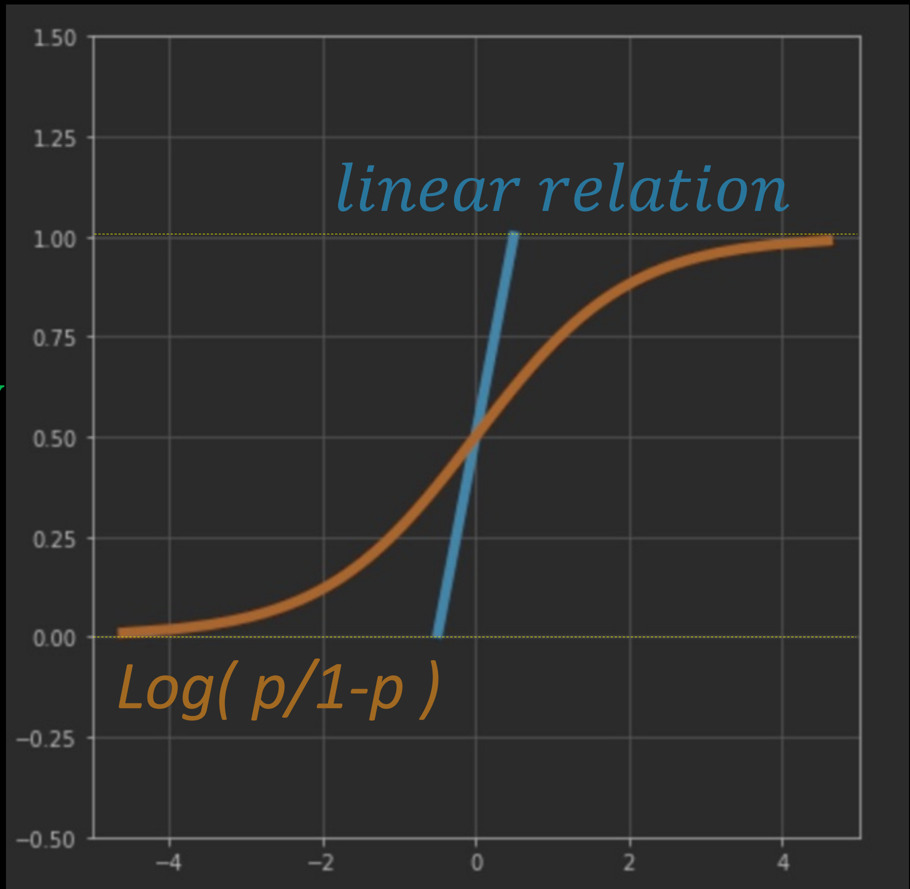

Clean cut
Clean cut
ambiguous



y
(Output Probability)
P=0.7
P=1.0
P=0.5
P=0.2
P=0
X
feature(s)


y
(Output Probability)
P=0.5
P>=0.5 Class=1
P<0.5 Class=0
X
feature(s)

Probability
0-1



feature $-\infty \sim \infty$

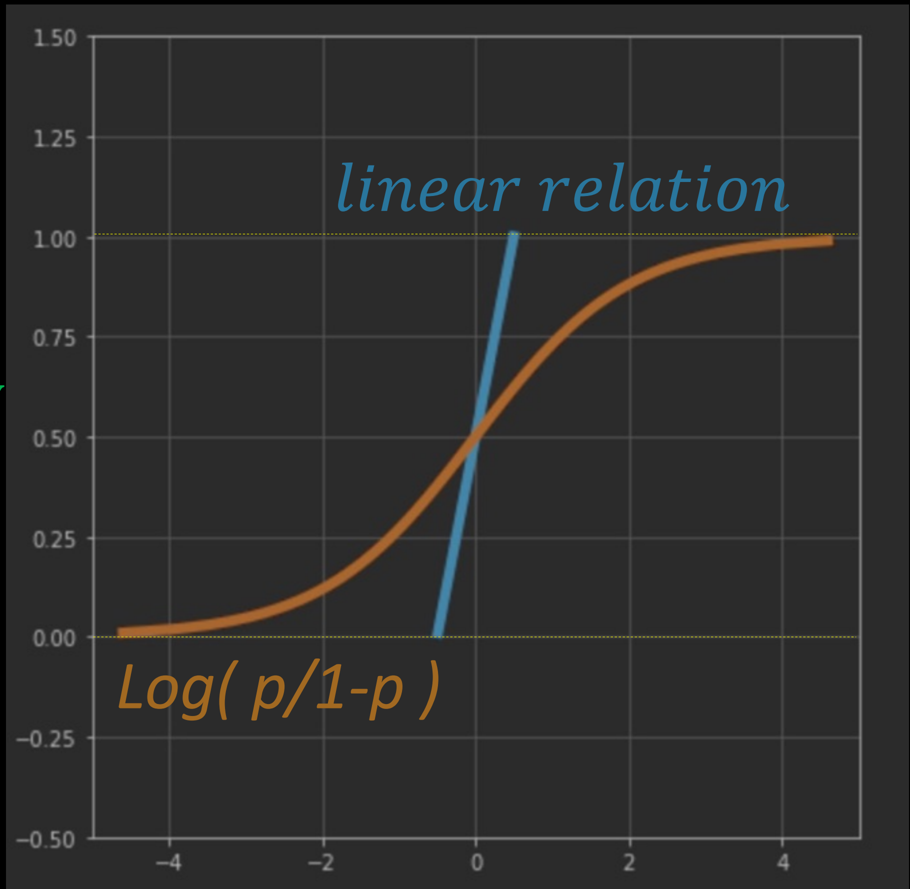
How to predict probability?

Remember we got output of range

$-\infty \sim \infty$

From regression

Probability
0-1



feature $-\infty \sim \infty$

Probability p $0 \sim 1$

Odd Ratio $\frac{p}{1-p}$ $0 \sim \infty$

Logit $\log(\frac{p}{1-p})$ $-\infty \sim \infty$

same range!

linear regression $\Sigma_i w_i x_i + b$ $-\infty \sim \infty$

Now we can just do regression to logit!

	No. of case patients	No. of control subjects	Odds ratio	95% confidence interval	*P* for trend*
Nonsmokers	117	1750	1.0	Referent	
Duration of tobacco use, y					
0.1–20.0	3	33	1.3	0.4–4.5	
20.1–32.0	7	33	3.4	1.4–8.0	
32.1–44.0	21	33	13.3	7.2–24.9	
⩾44.1	30	30	19.1	10.4–35.1	<.0001
Average consumption of tobacco, g/day					
0.1–3.5	2	10	2.2	0.5–10.4	
3.6–5.0	22	54	7.9	4.3–14.3	
5.1–10.7	6	18	4.8	1.9–12.6	
⩾10.8	31	47	12.4	7.2–21.4	.1
Cumulative consumption of tobacco, g/day × y					
1–71	3	31	1.3	0.4–4.4	
72–157	13	34	7.6	3.8–15.4	
158–382	15	32	8.0	4.0–15.7	
⩾383	30	32	18.3	10.2–32.8	.0006
Age at start of tobacco use, y					
⩽19	27	48	9.6	5.6–16.7	
20–26	20	52	6.3	3.5–11.2	
⩾27	14	29	8.2	4.1–16.3	.4

*Test for linear trend, two-sided *P* value (considered statistically significant for $P<.05$).

Now we are mapping the output of regression:

To logit

$\Sigma_i w_i x_i + b$ $-\infty \sim \infty$ to $Logit$ $\log(\frac{p}{1-p})$ $-\infty \sim \infty$

$$p = \frac{1}{1+exp-(\Sigma_i w_i x_i + b)} = \text{Sigmoid}(w_i x_i + b) = S(\beta X)$$

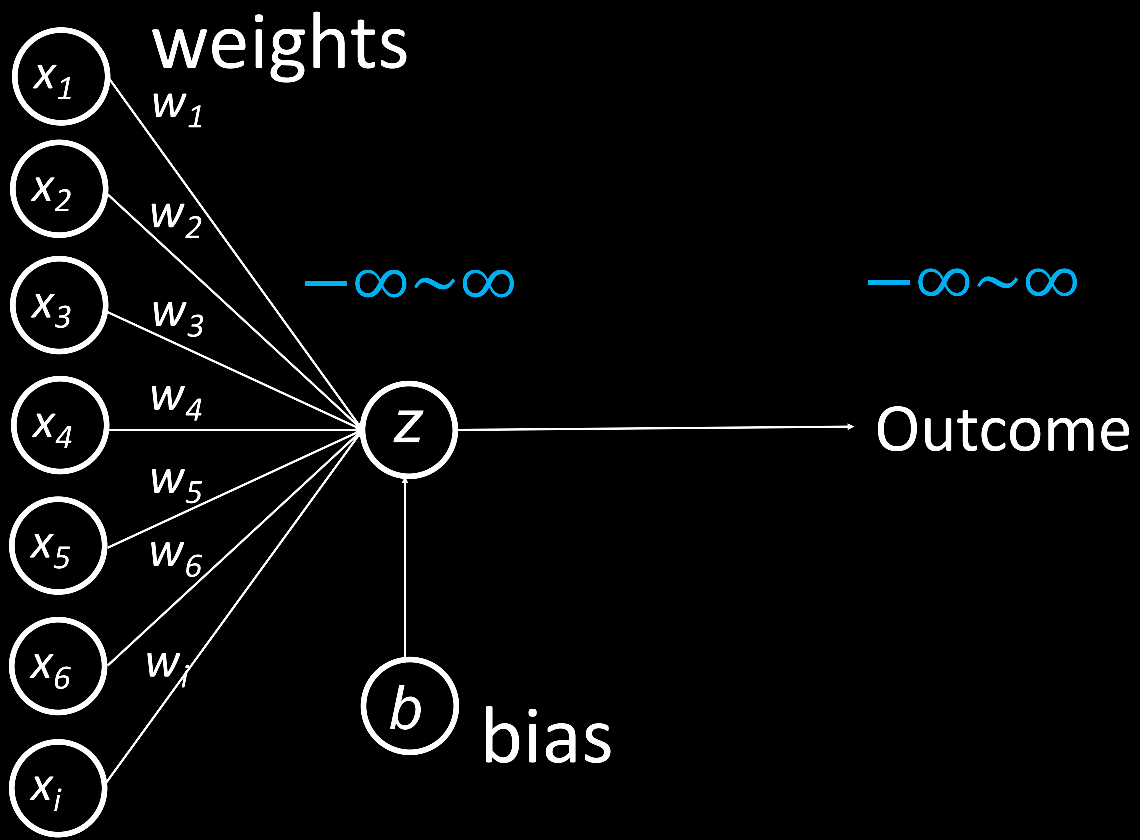
Just
Logistic
Regression

$$p = \frac{1}{1+exp-(\Sigma_i w_i x_i + b)} = \text{Sigmoid}(w_i x_i + b) = S(\beta X)$$

$P = S(\beta X)$

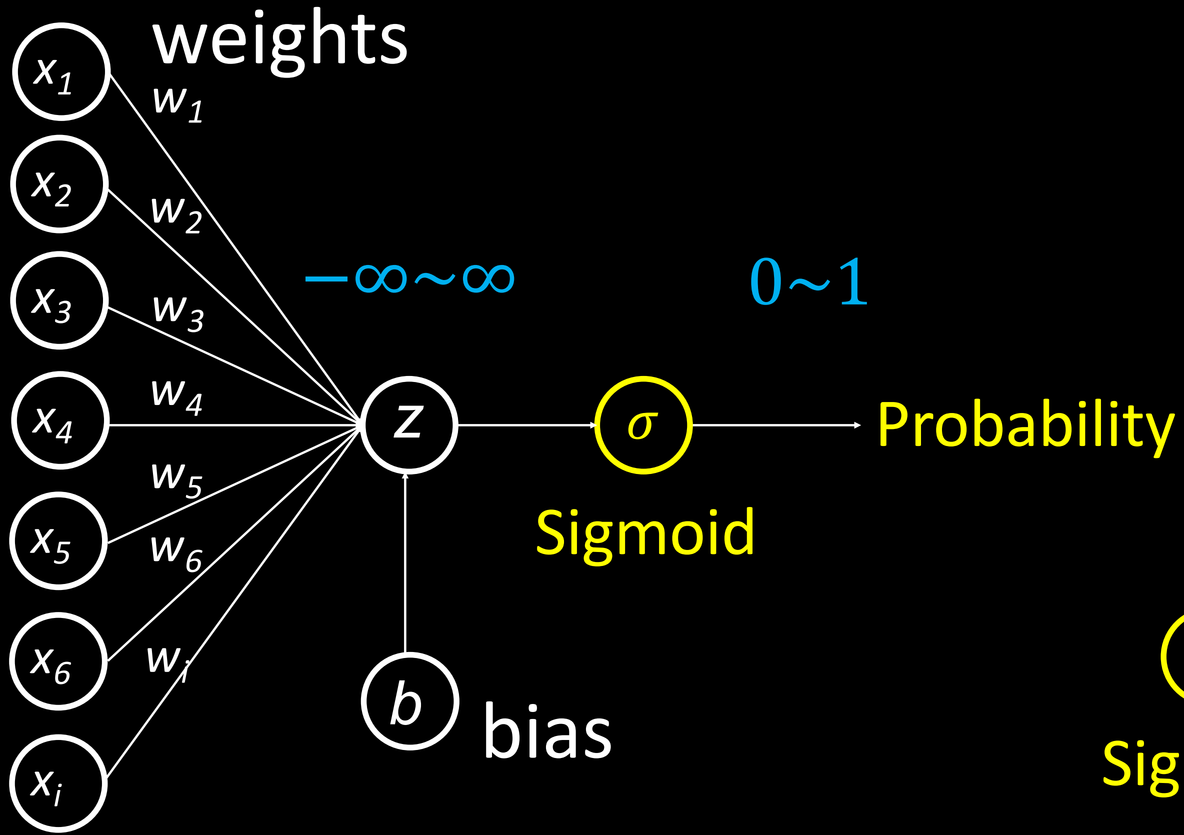
Logistic Regression = Regression
Between
Probability and Sigmoid of
(weighted features)

Graph of Linear Regression



Graph of Logistic Regression

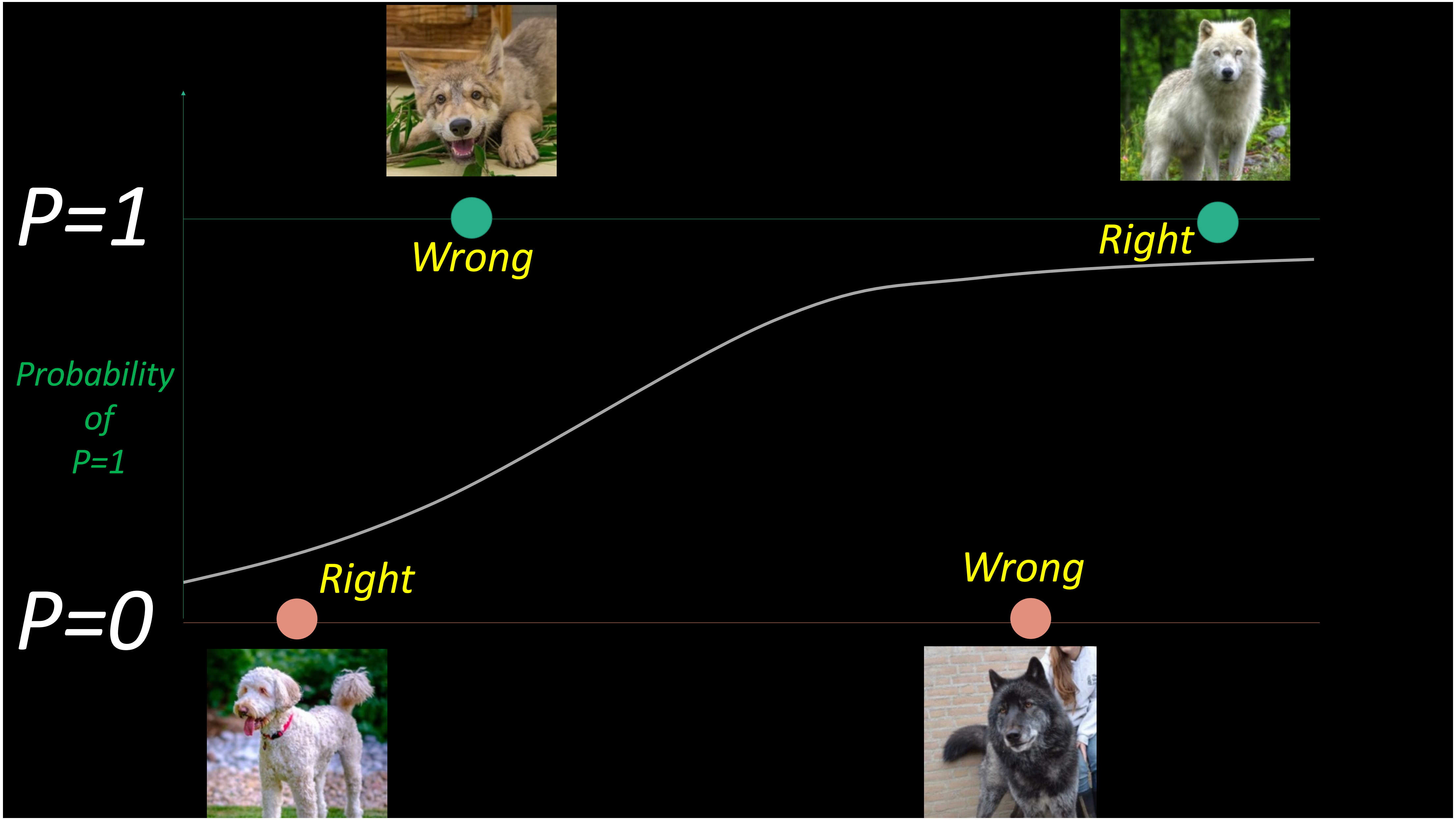
$$p = \frac{1}{1+exp-(\Sigma_i w_i x_i + b)} = \text{Sigmoid}(\Sigma_i w_i x_i + b) = S(WX)$$



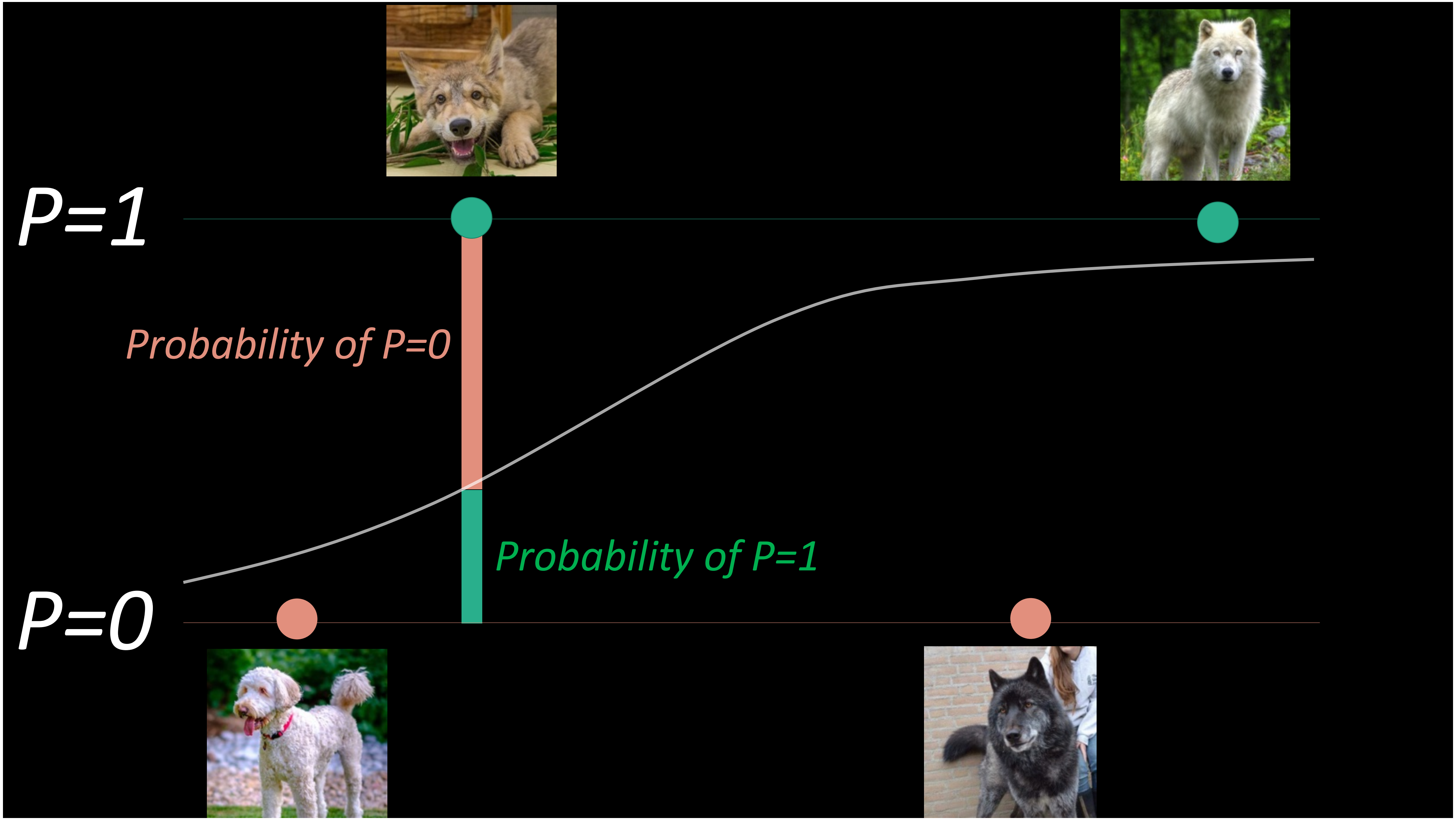
Just add 1 more "layer"

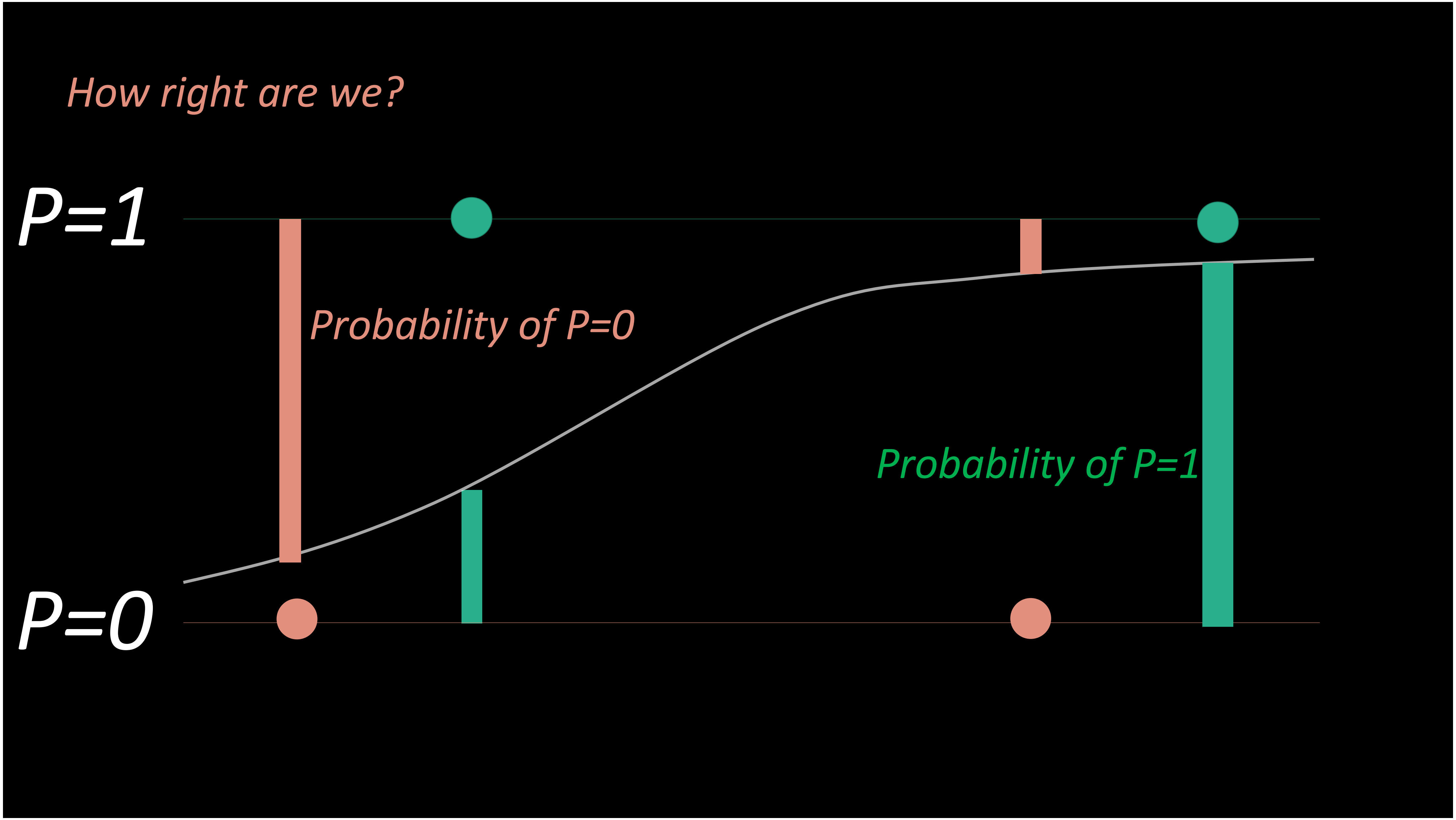
Loss Function for Classification

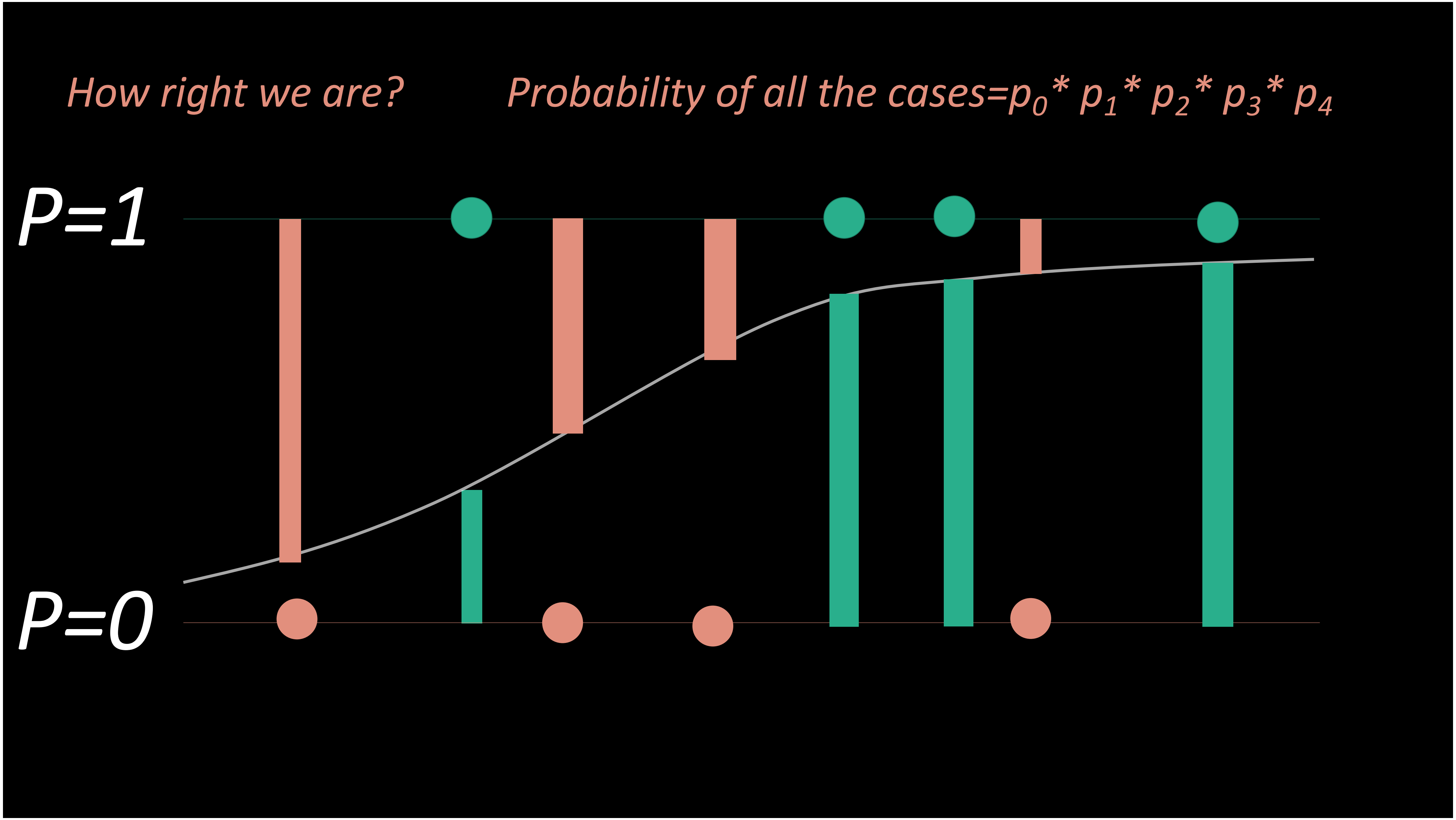
(need to way to say good or bad)



P=1
Wrong
Right
Probability
of
P=1
Right
Wrong
P=0


P=1
Probability of P=0
Probability of P=1
P=0


How right are we?
P=1
Probability of P=0
Probability of P=1
P=0


How right we are?
Probability of all the cases=p_0* p_1* p_2* p_3* p_4
P=1
P=0

Probability of all the cases

$$(+)\; p_0 * p_1 * p_2 * \dots\dots$$

$$(-)\; p_0 * p_1 * p_2 * \dots\dots$$

Remember log(Prob.)

$$\prod_n P(y_n | x_n; \theta)$$

VS

$$\sum_i logP(y_n | x_n; \theta)$$

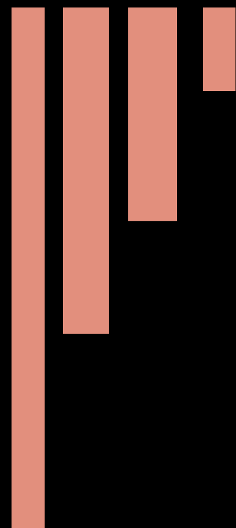
Binary Cross Entropy (BCE) loss

y = 1   *y = 0*

$$BCE = -\sum_n \overset{1}{y} \cdot \log(p(\hat{y})) \quad -\sum_n \overset{1}{(1-y)} \cdot \log(p(1-\hat{y}))$$

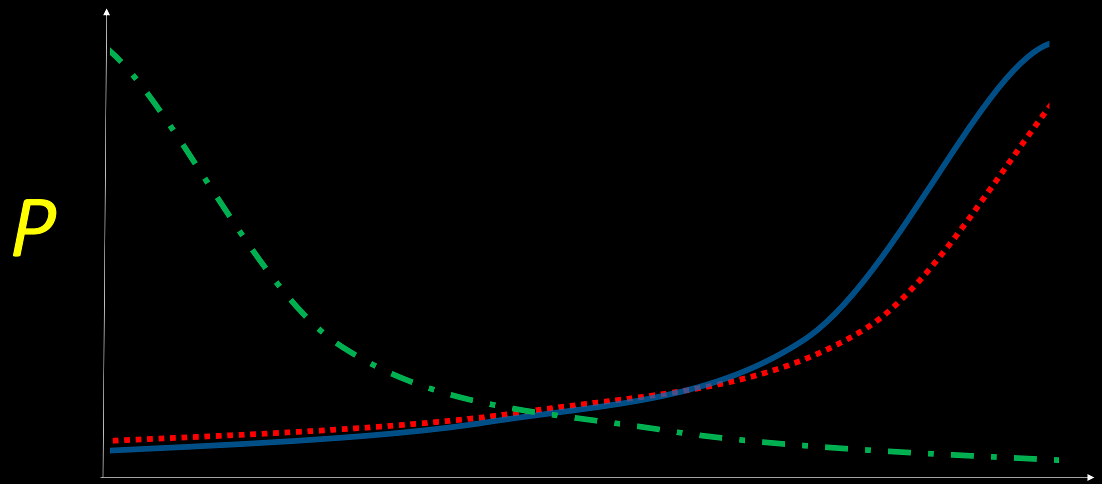
All the probs predicted y = 1 correctly

All the probs predicted y = 0 correctly



Again, this is based on Maximum Likelihood Estimation (MLE)

$$\theta_{ML} = argmax_{\theta} P(y|X;\theta)$$



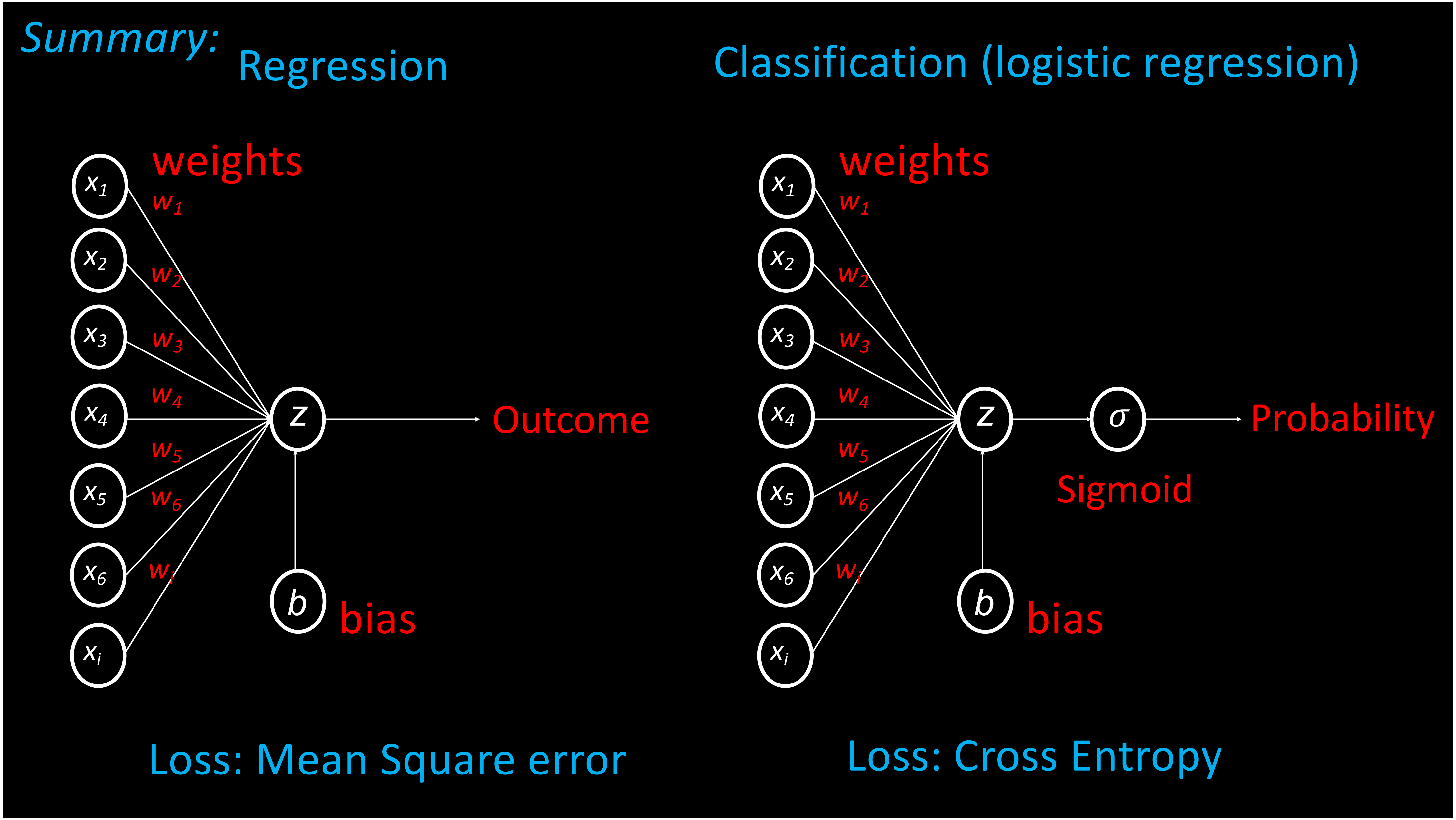
Q, Real Probability

$P(y|X;\theta)$, good prediction

$P(y|X;\theta)$, bad prediction

Cross Entropy:

how different they are?



Summary:
Regression
Classification (logistic regression)
weights
x_1 x_2 x_3 x_4 x_5 x_6 x_i
w_1 w_2 w_3 w_4 w_5 w_6 w_i
z
Outcome
b
bias
weights
x_1 x_2 x_3 x_4 x_5 x_6 x_i
w_1 w_2 w_3 w_4 w_5 w_6 w_i
z
σ
Probability
Sigmoid
b
bias
Loss: Mean Square error
Loss: Cross Entropy