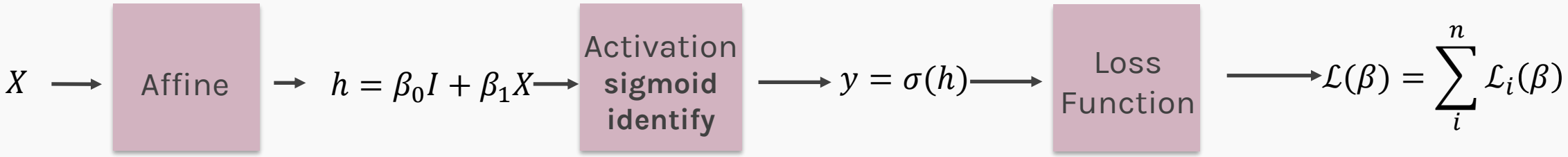
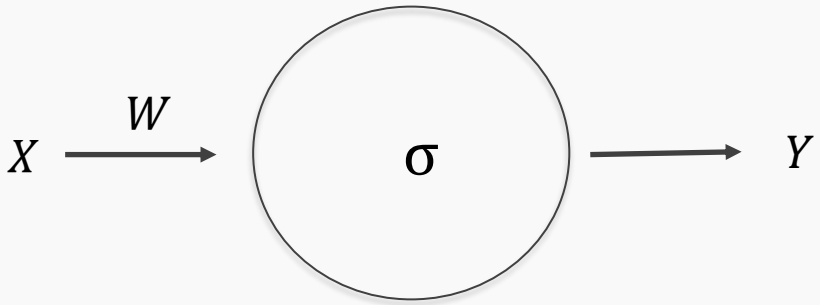
Multilayer Perceptron

Pavlos Protopapas

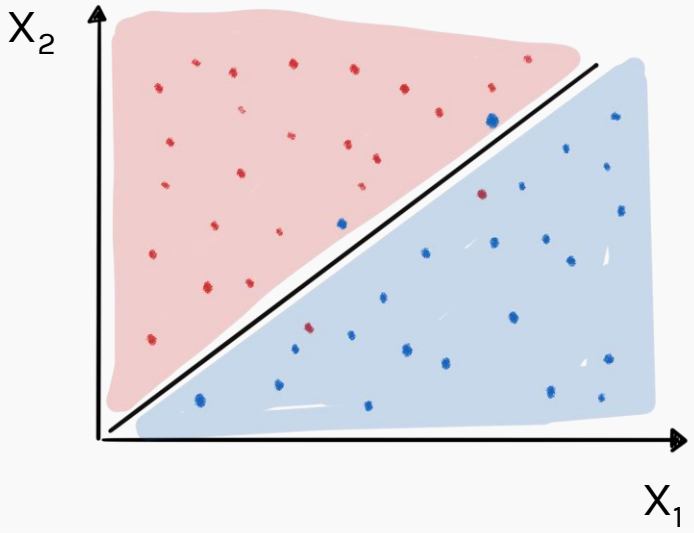
A single neuron

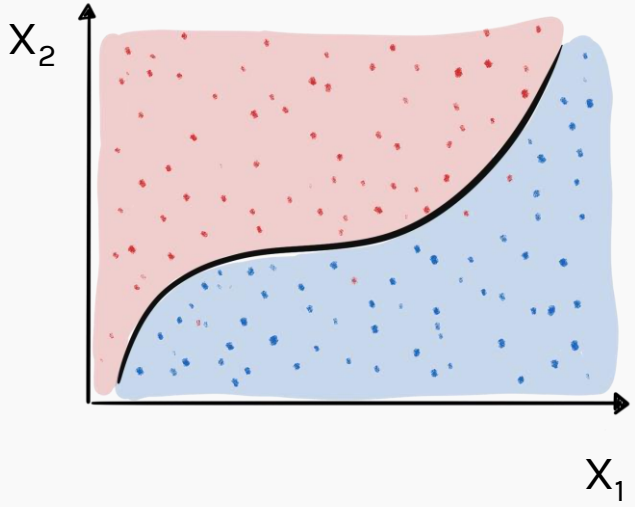
Up to this point we just re-branded logistic regression or linear regression to look like a neuron.

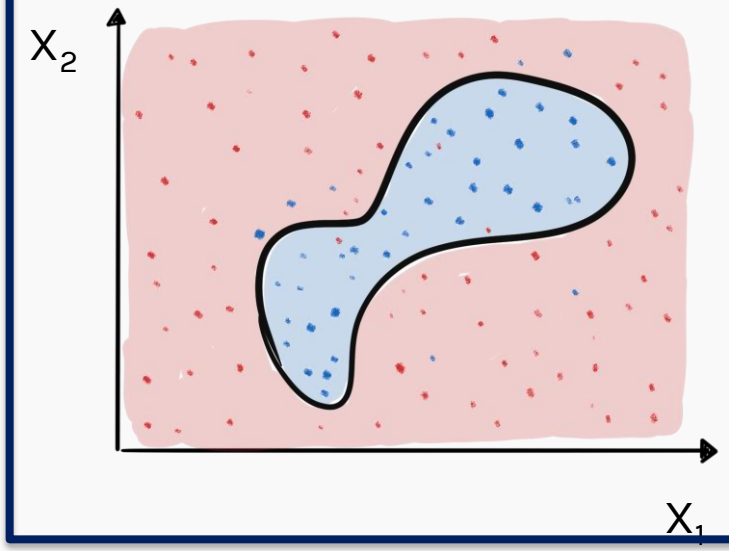


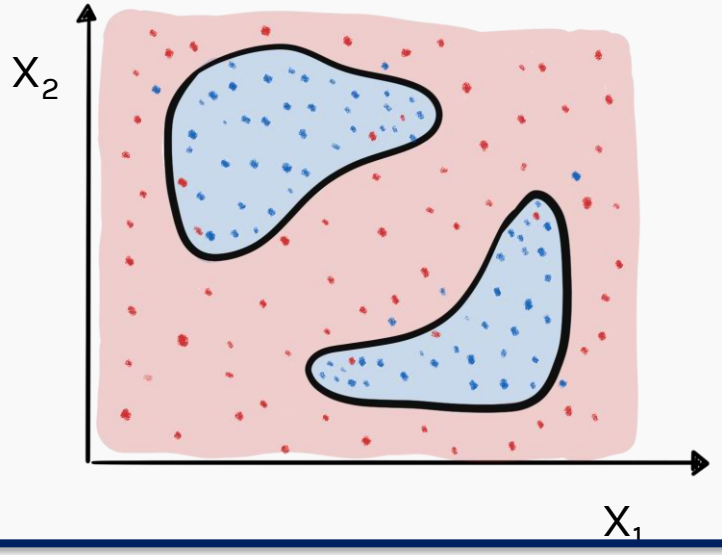


So, what's the big deal about Neural Networks?



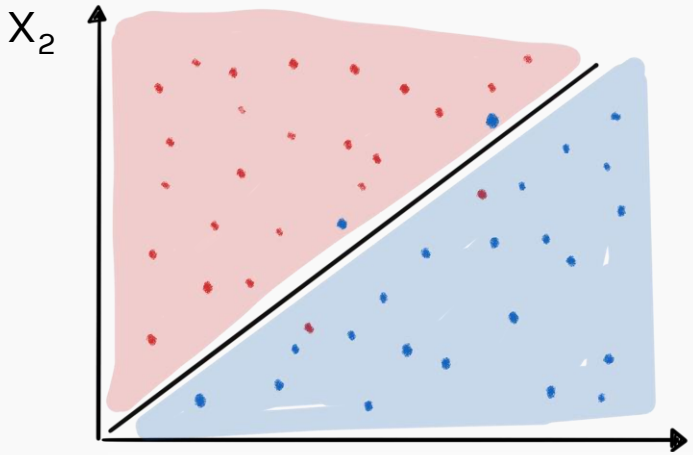


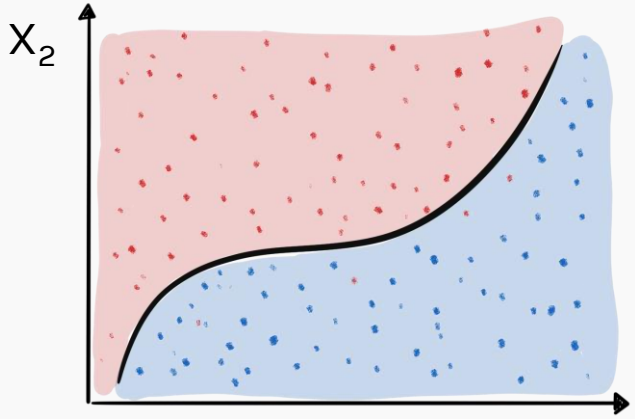




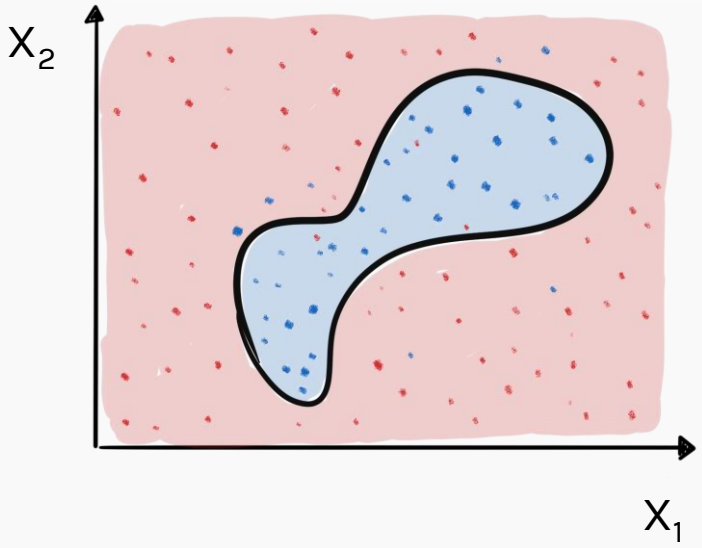
Very hard to express boundaries like these with logistic regression

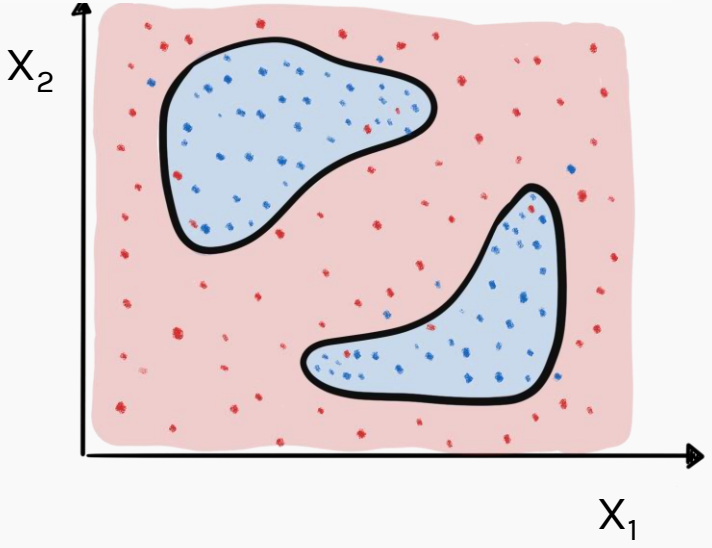
So, what's the big deal about Neural Networks?



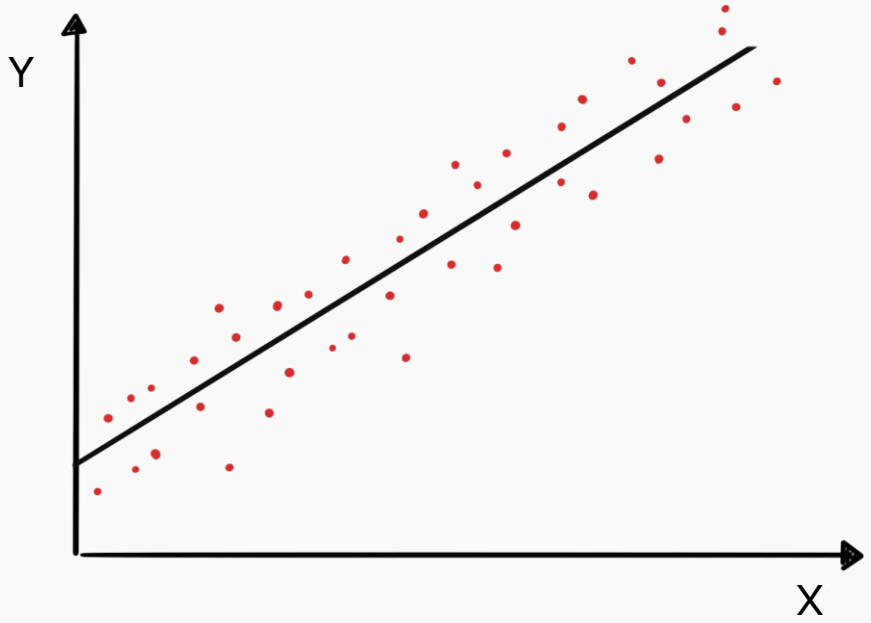


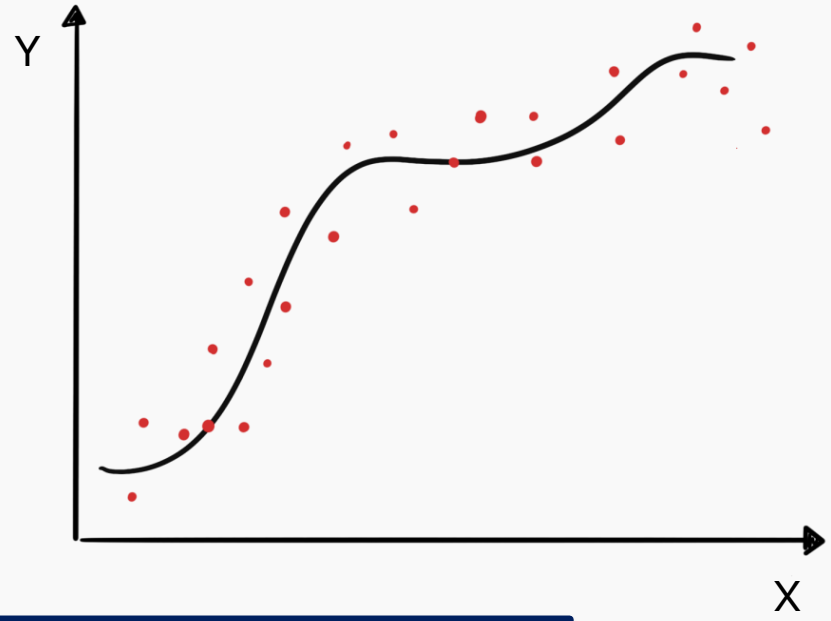
No Linear Boundaries!

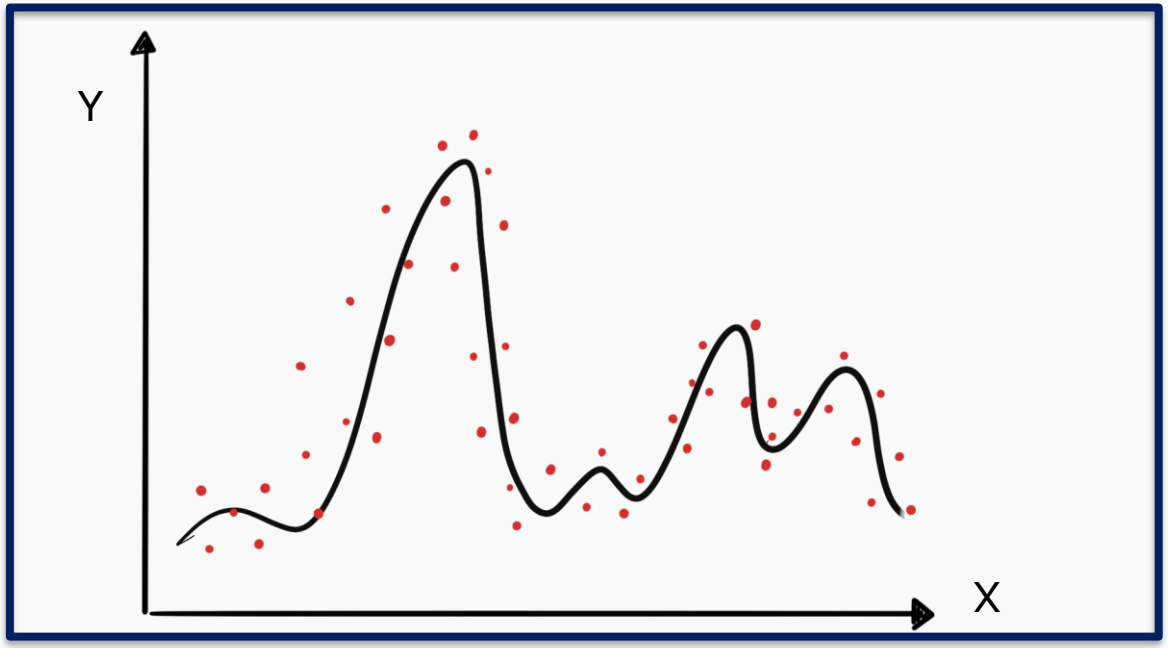




So, what's the big deal about Neural Networks?

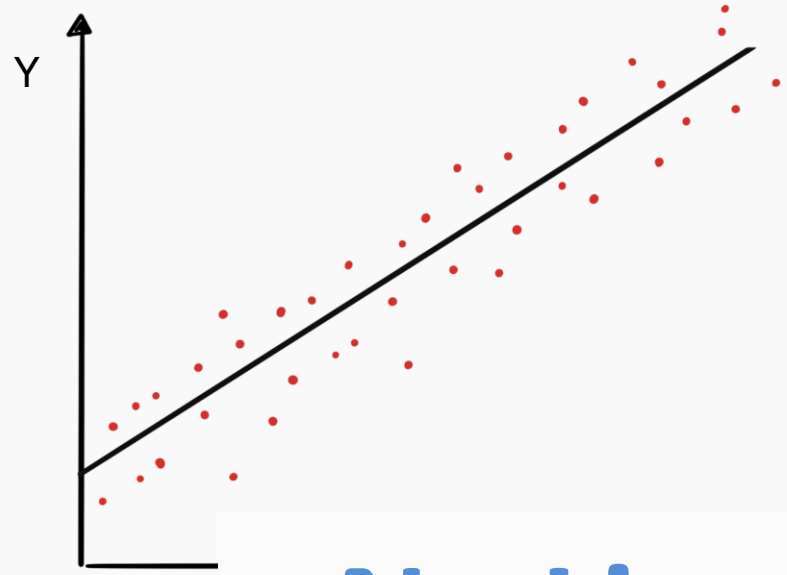


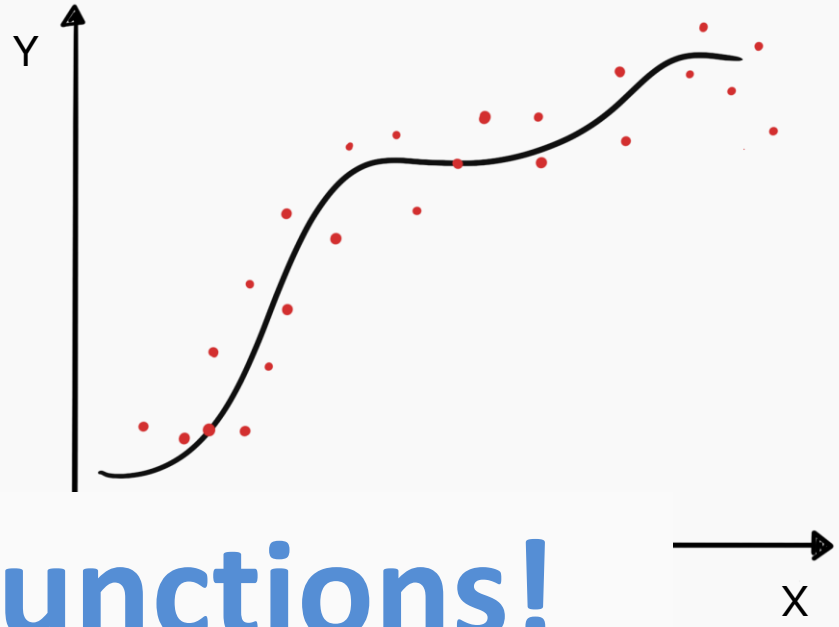




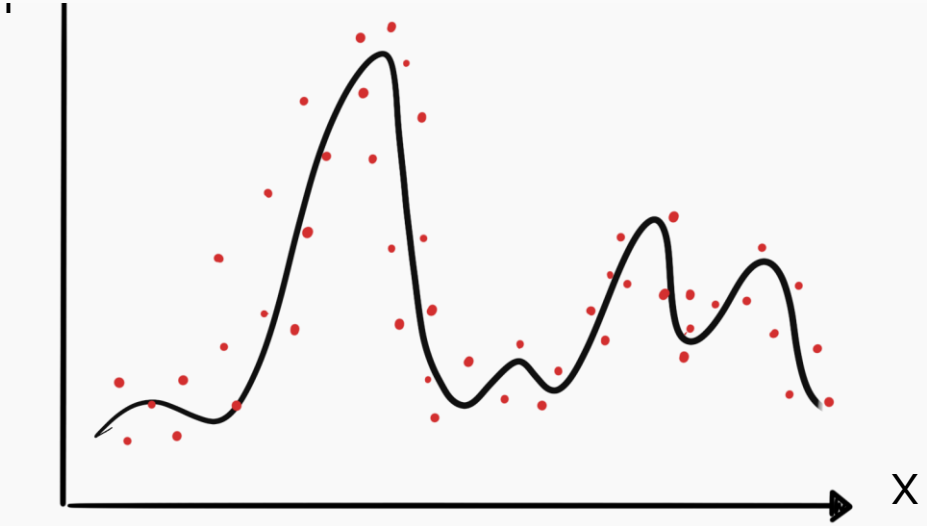
Very hard to express functions like these with polynomial regression.

So, what's the big deal about Neural Networks?



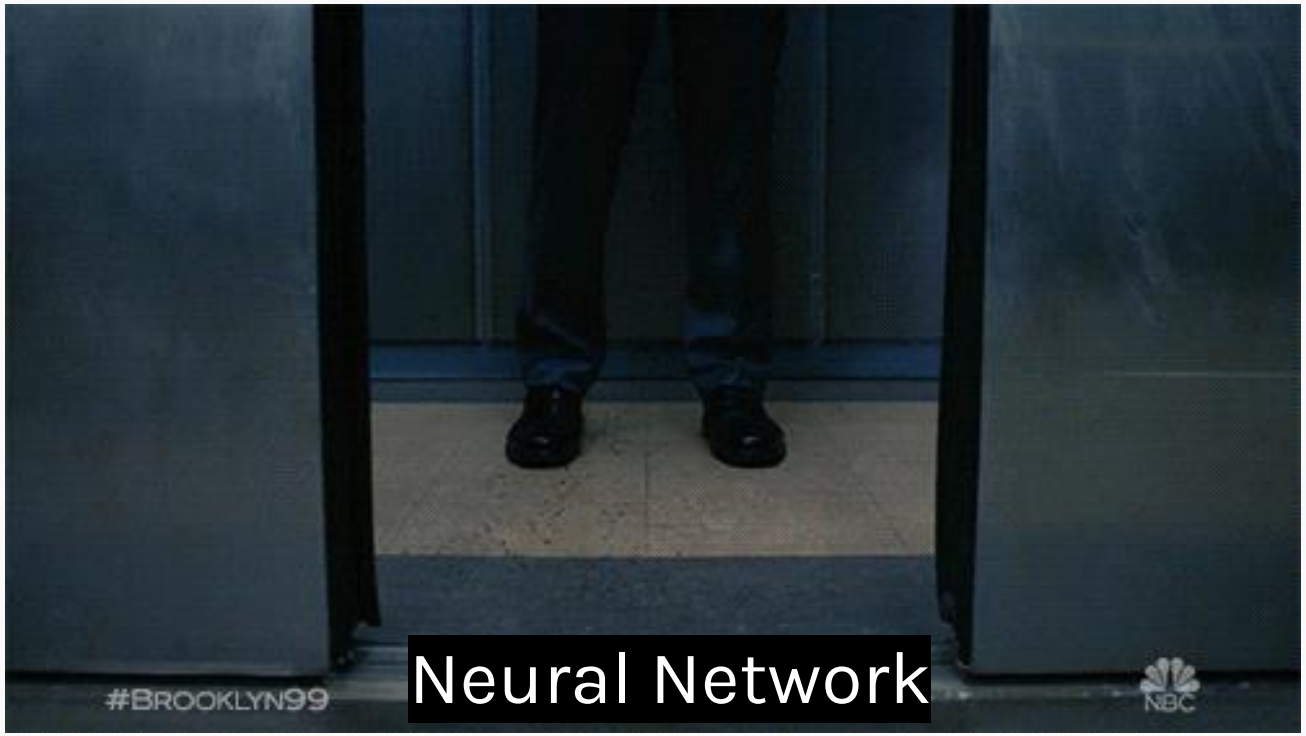


No Linear Functions!



So, what's the big deal about Neural Networks?

How can we overcome these challenges?



Outline

- Introduction to Artificial Neural Networks
- Review of Classification and Logistic Regression
- Single Neuron Network ('Perceptron')
- **Multi-Layer Perceptron (MLP)**

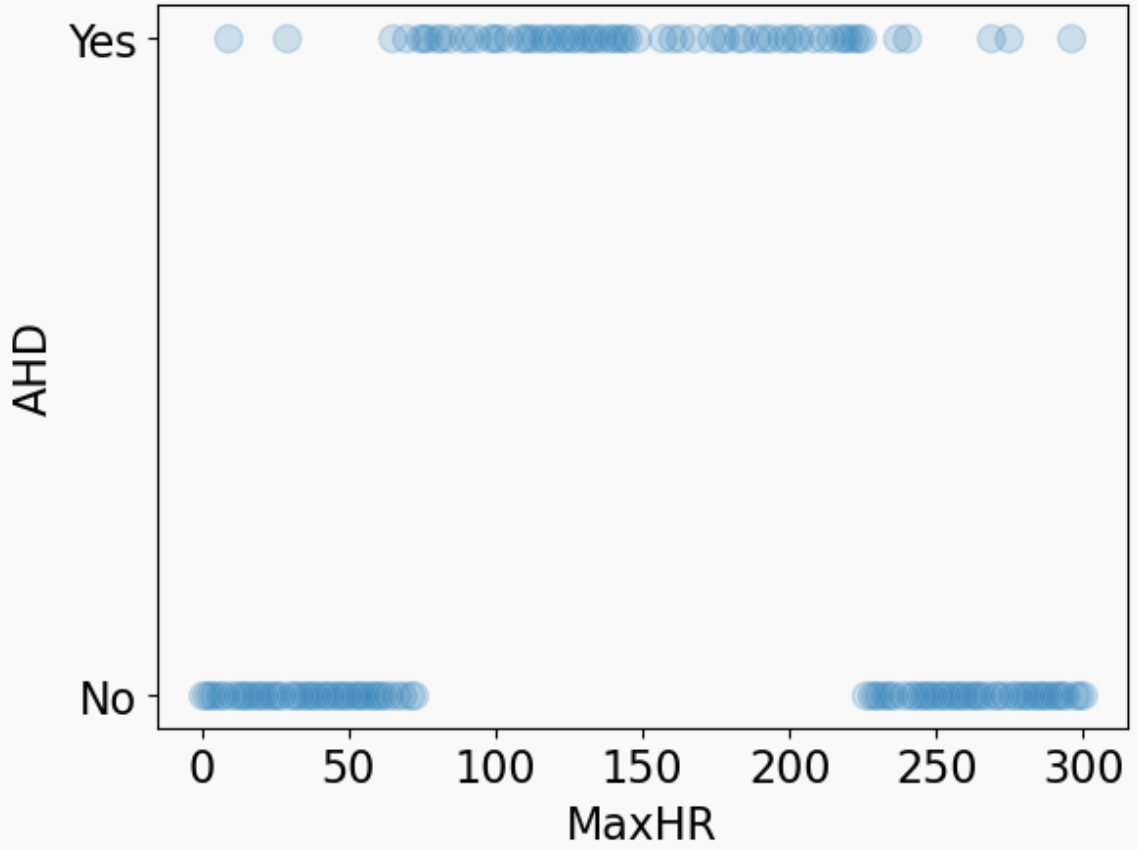
Example Using Heart Data

Consider a dataset that contains a binary outcome AHD for 303 patients who presented with chest pain.

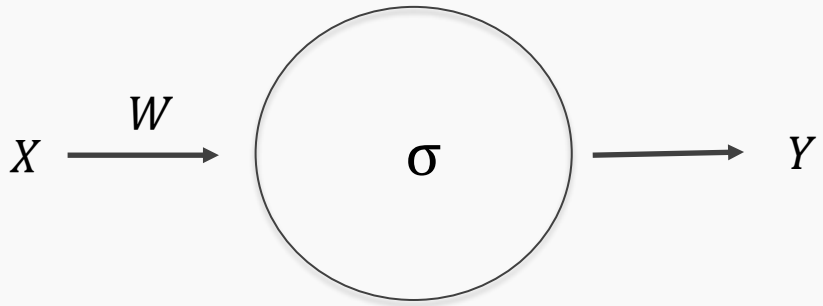
Age	Sex	ChestPain	RestBP	Chol	MaxHR	ExAng	Thal	AHD
63	1	typical	145	233	150	0	fixed	No
67	1	asymptomatic	160	286	108	1	normal	Yes
67	1	asymptomatic	120	229	129	1	reversable	Yes
37	1	nonanginal	130	250	187	0	normal	No

Example Using Heart Data

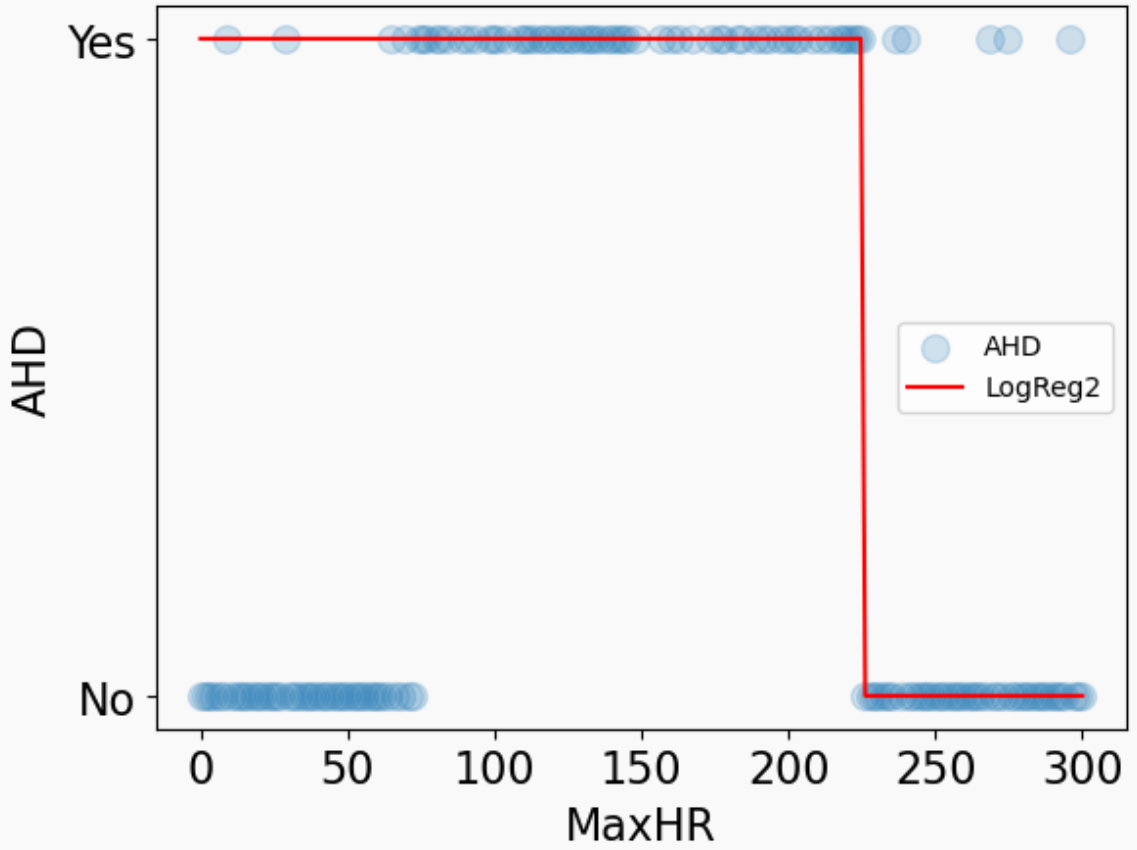
Slightly modified data to illustrate concepts.



Example Using Heart Data



Choose W such as **right** part of data is fitted well.



Quiz Time

What would be the expected behavior of the logistic regression function below if we increase the bias (b or W_0)?

A. The curve will shift upwards

B. The curve will shift downwards

C. The curve will shift to the left

D. The curve's slope will become steeper

Quiz Time

What would be the expected behavior of the logistic regression function below if we increase the bias (b or W_0)?

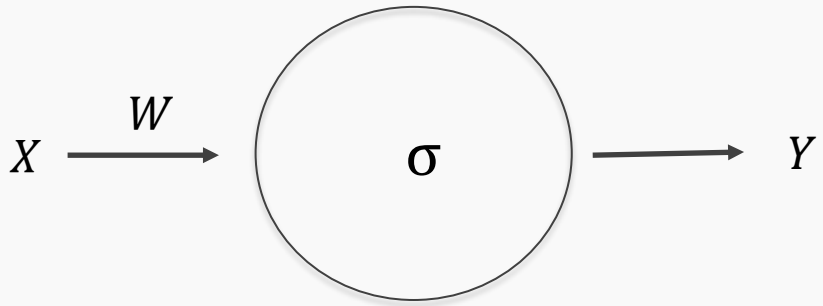
A. The curve will shift upwards

B. The curve will shift downwards

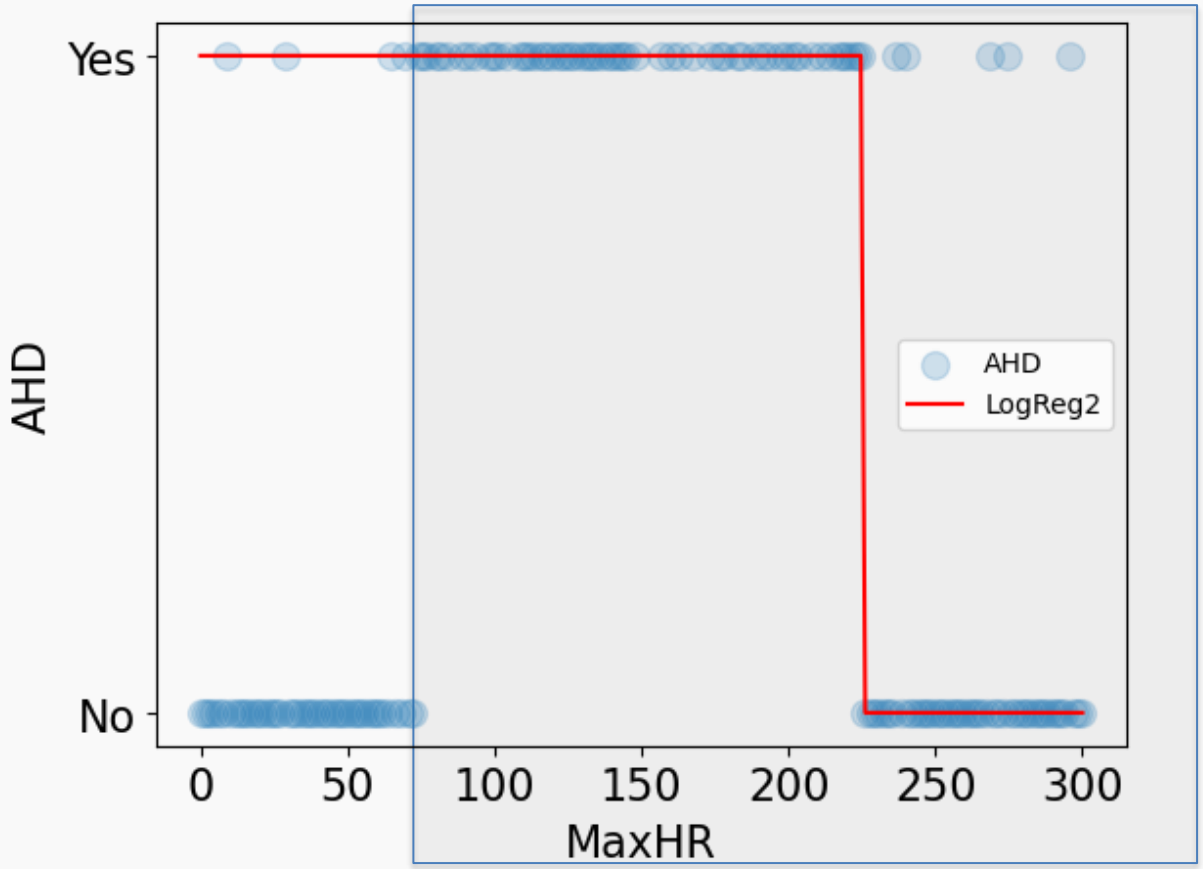
C. The curve will shift to the left

D. The curve's slope will become steeper

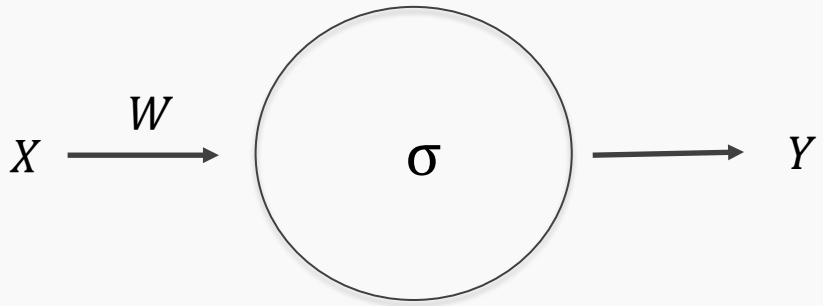
Example Using Heart Data



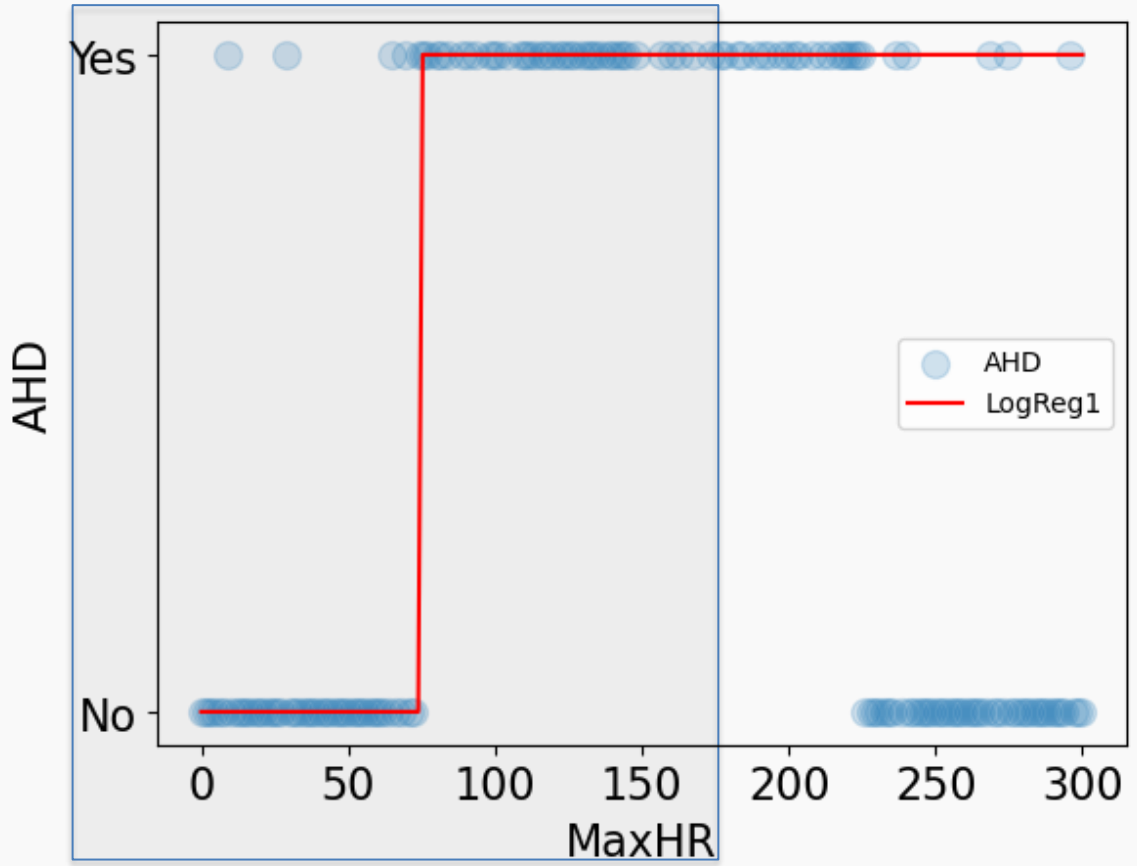
Choose W such as **right** part of data is fitted well



Example Using Heart Data

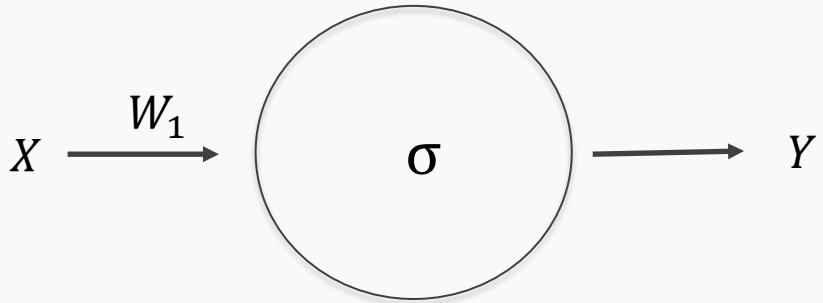


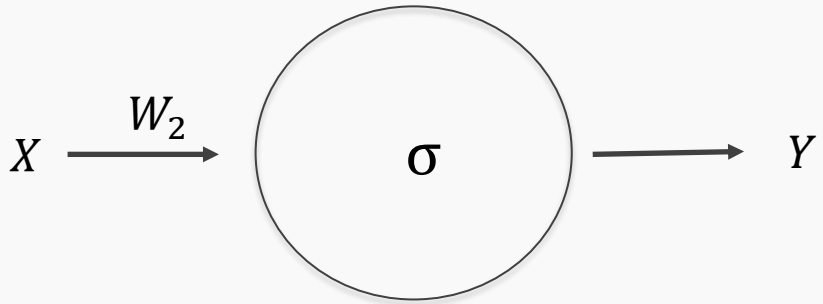
Choose W such as **left** part of data is fitted well

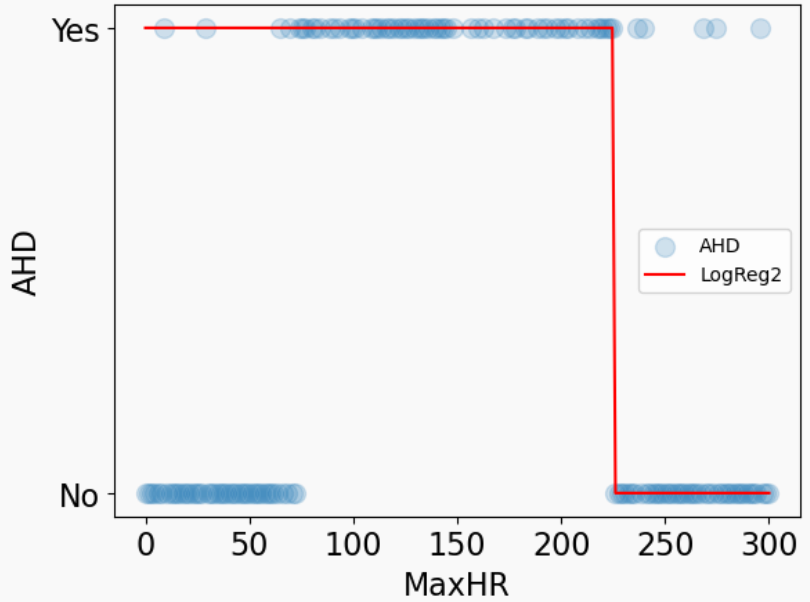


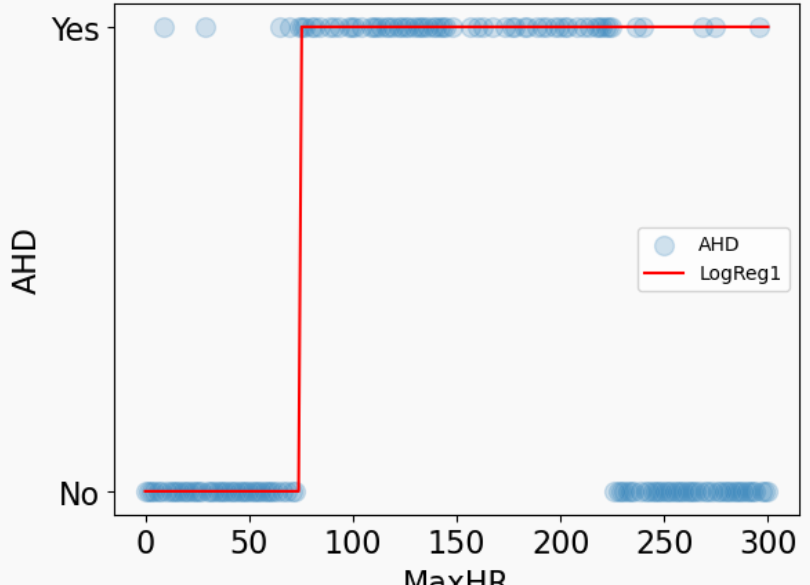
Example Using Heart Data

Two regions, two nodes

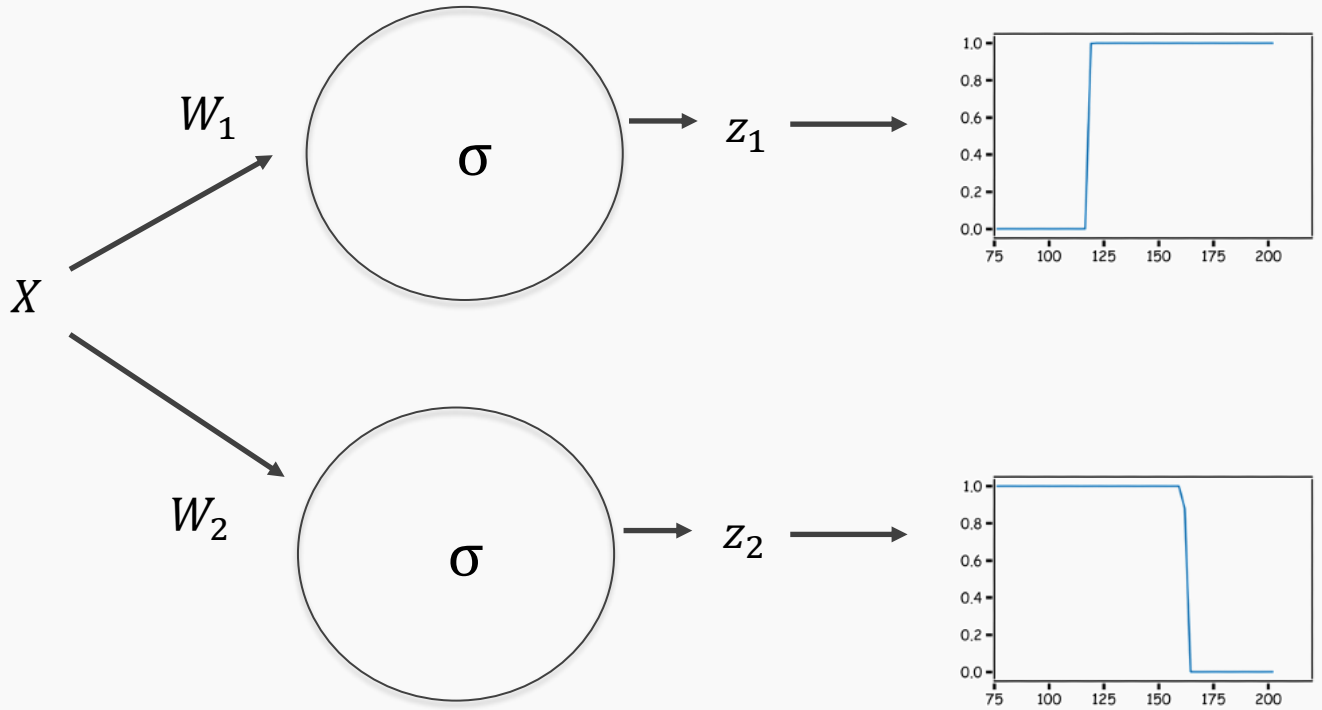








Combining Neurons



Quiz Time

How shall we combine these results?

A. Multiply z_1 and z_2

B. Add z_1 and z_2

C. $\sigma(z_1) + \sigma(z_2)$

D. Convolve z_1 with z_2

Quiz Time

How shall we combine these results?

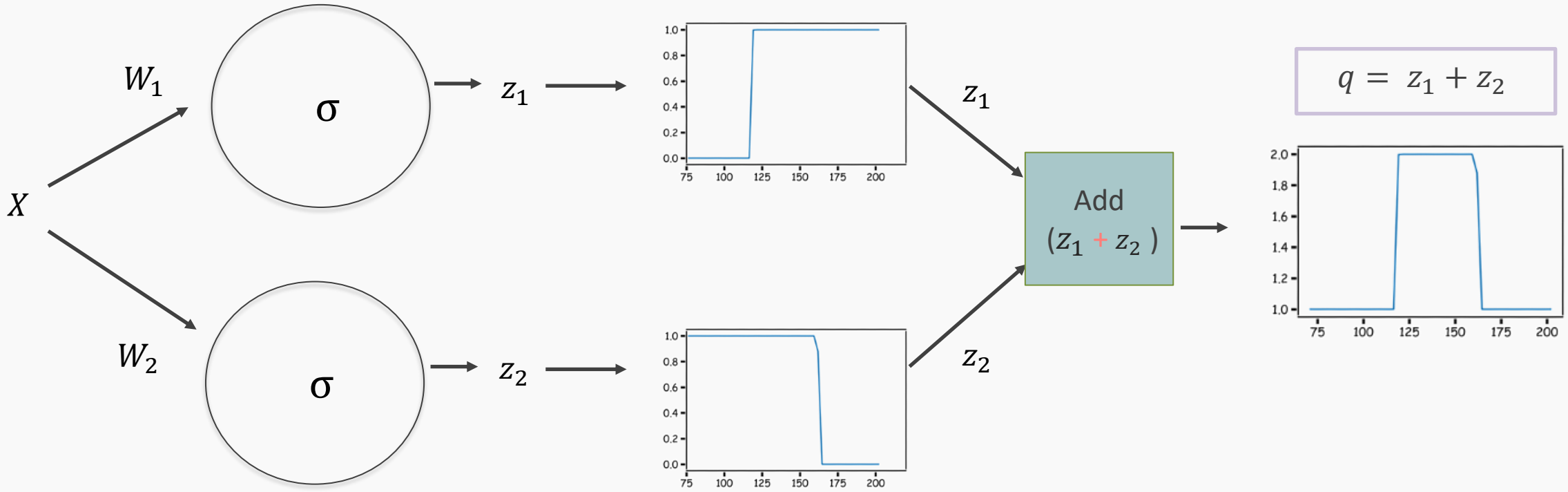
A. Multiply z_1 and z_2

B. Add z_1 and z_2

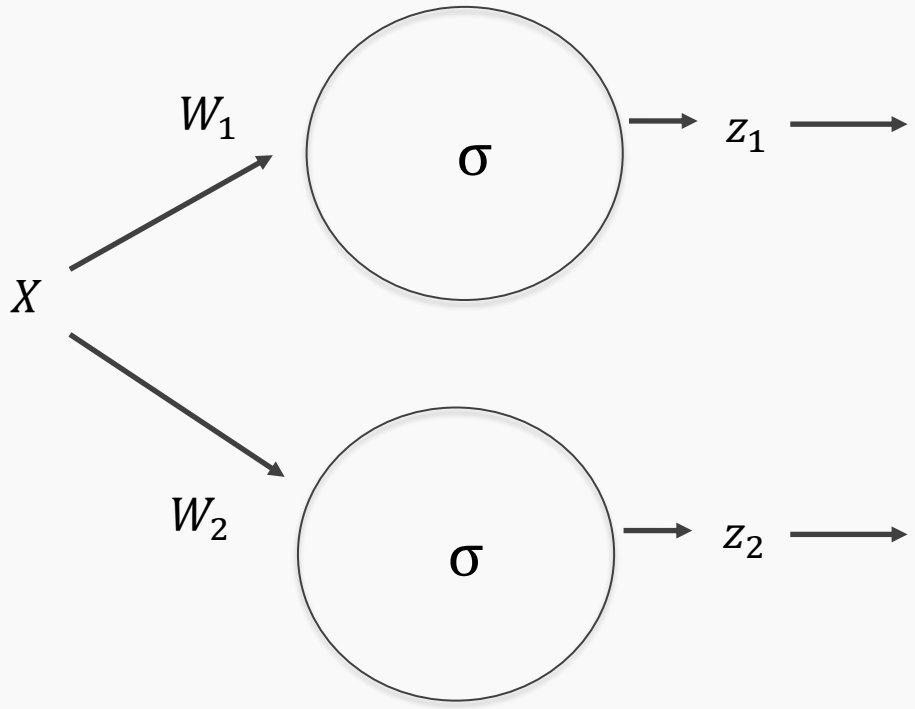
C. $\sigma(z_1) + \sigma(z_2)$

D. Convolve z_1 with z_2

Combining Neurons

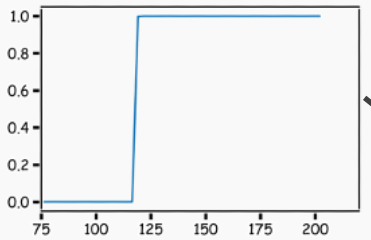


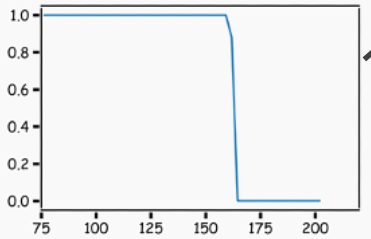
Combining Neurons



z_1

z_2



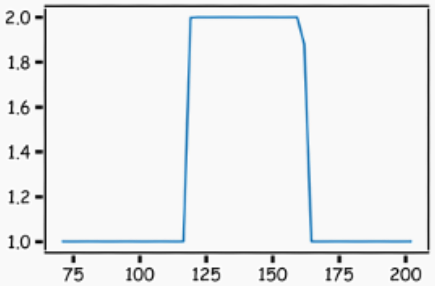


z_1

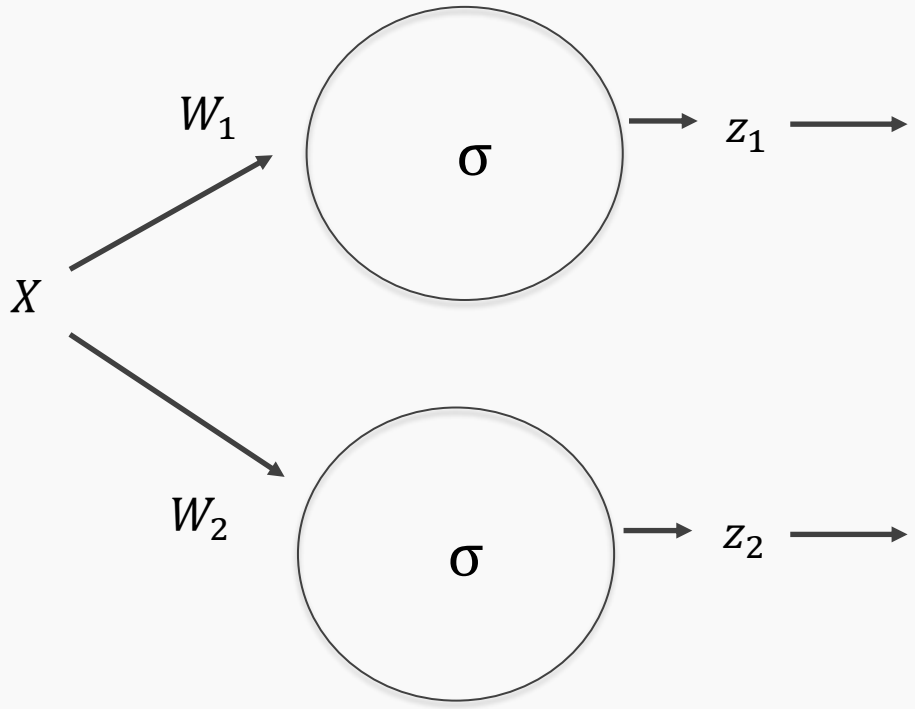
z_2

Affine

$$q = W_{31}z_1 + W_{32}z_2 + W_{30}$$

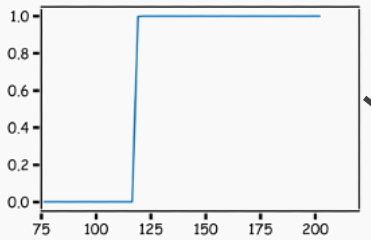


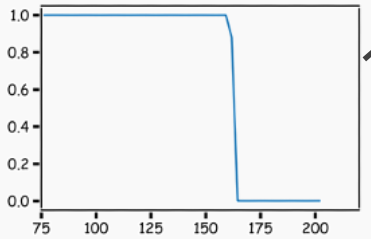
Combining Neurons



z_1

z_2



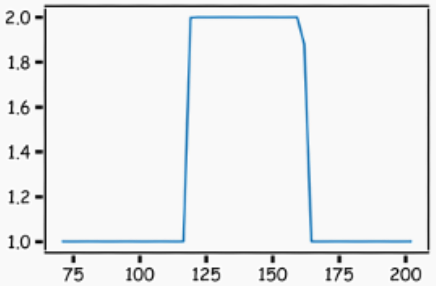


z_1

z_2

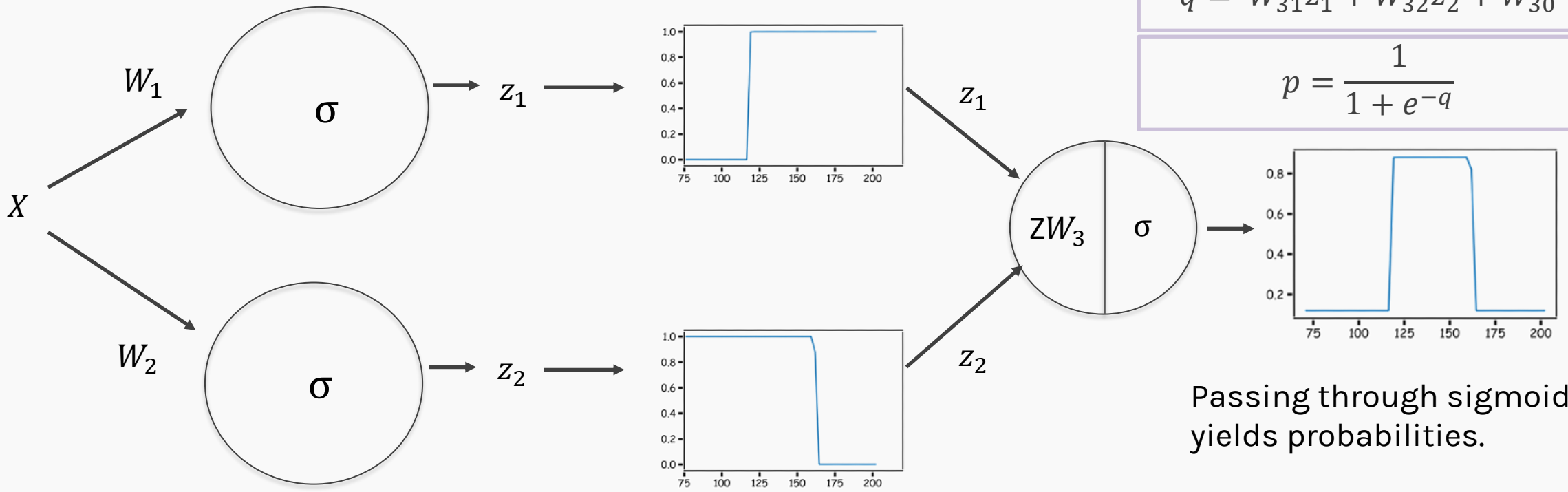
Affine

$$q = W_{31}z_1 + W_{32}z_2 + W_{30}$$



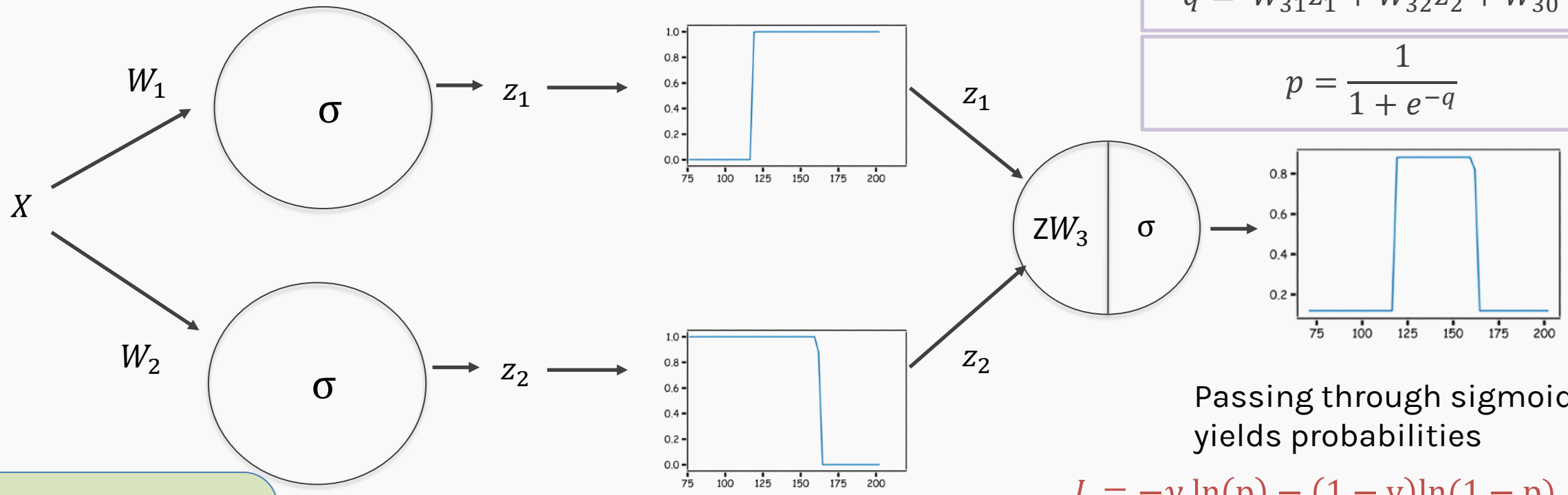
But this is not a probability.

Combining Neurons



Passing through sigmoid yields probabilities.

Combining Neurons



$$q = W_{31}z_1 + W_{32}z_2 + W_{30}$$

$$p = \frac{1}{1 + e^{-q}}$$

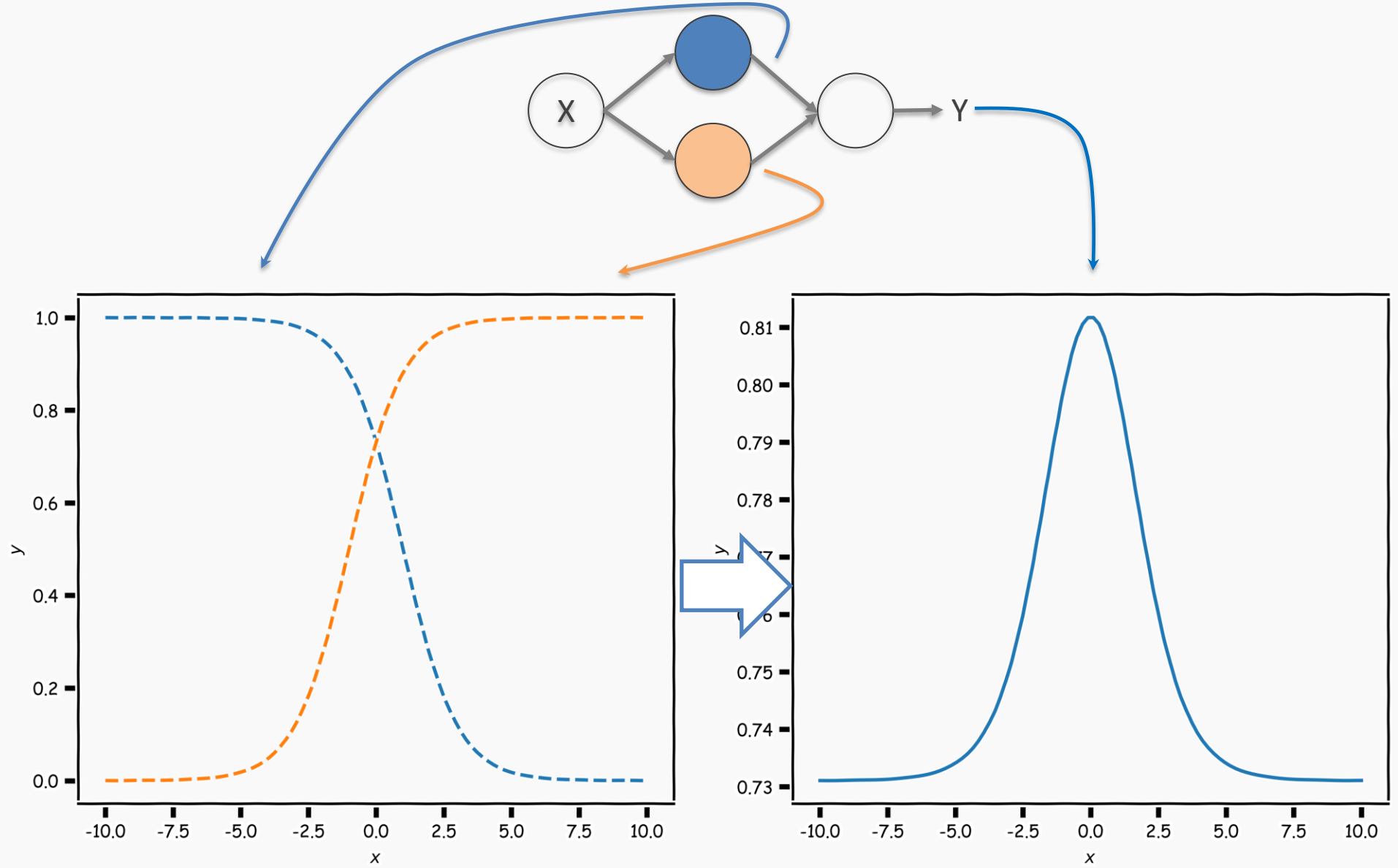
Passing through sigmoid yields probabilities

$$L = -y\ln(p) - (1-y)\ln(1-p)$$

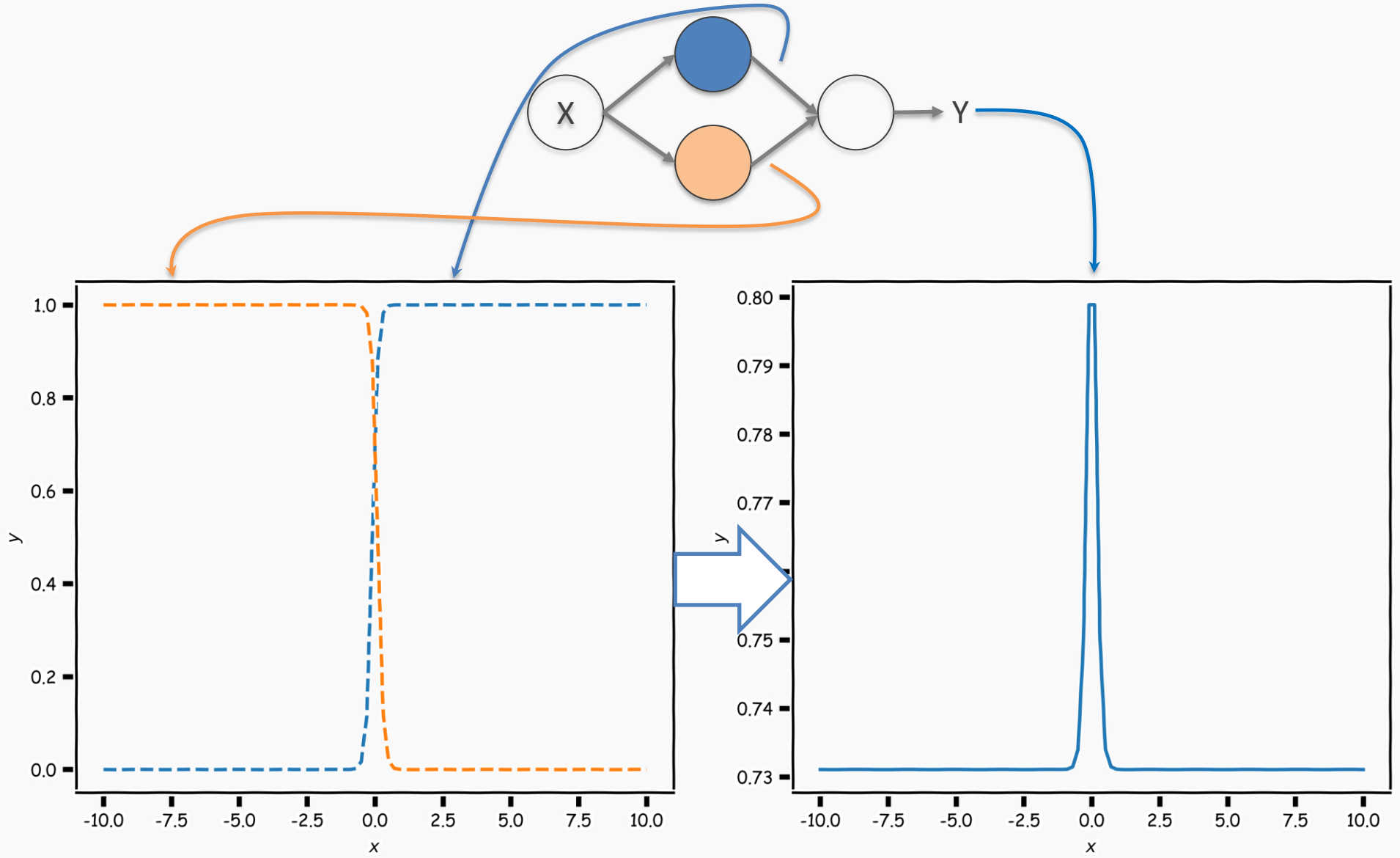
We will talk about about this in the coming lectures

Need to learn W_1 ,W_2 and W_3

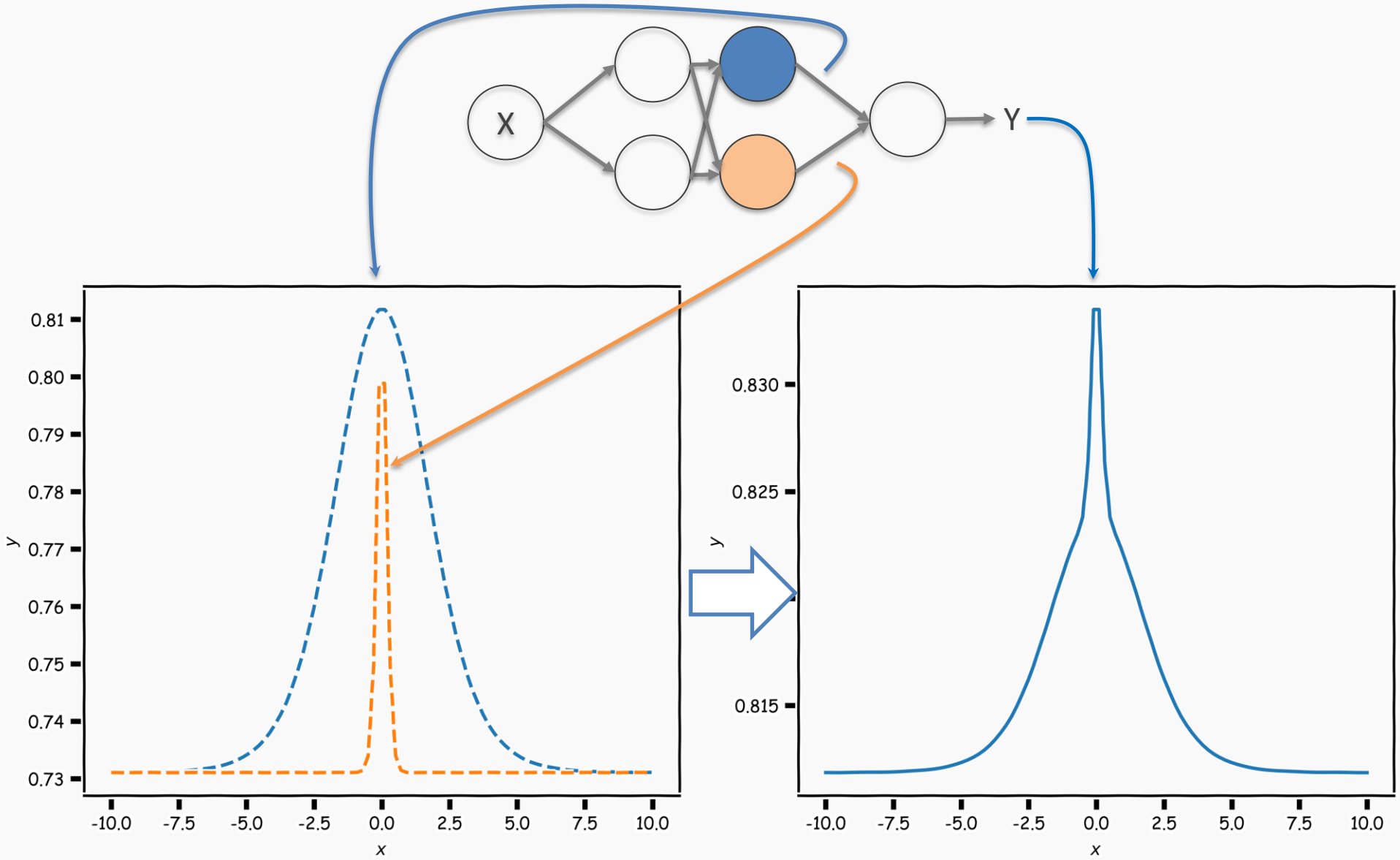
Combining neurons allows us to model interesting functions



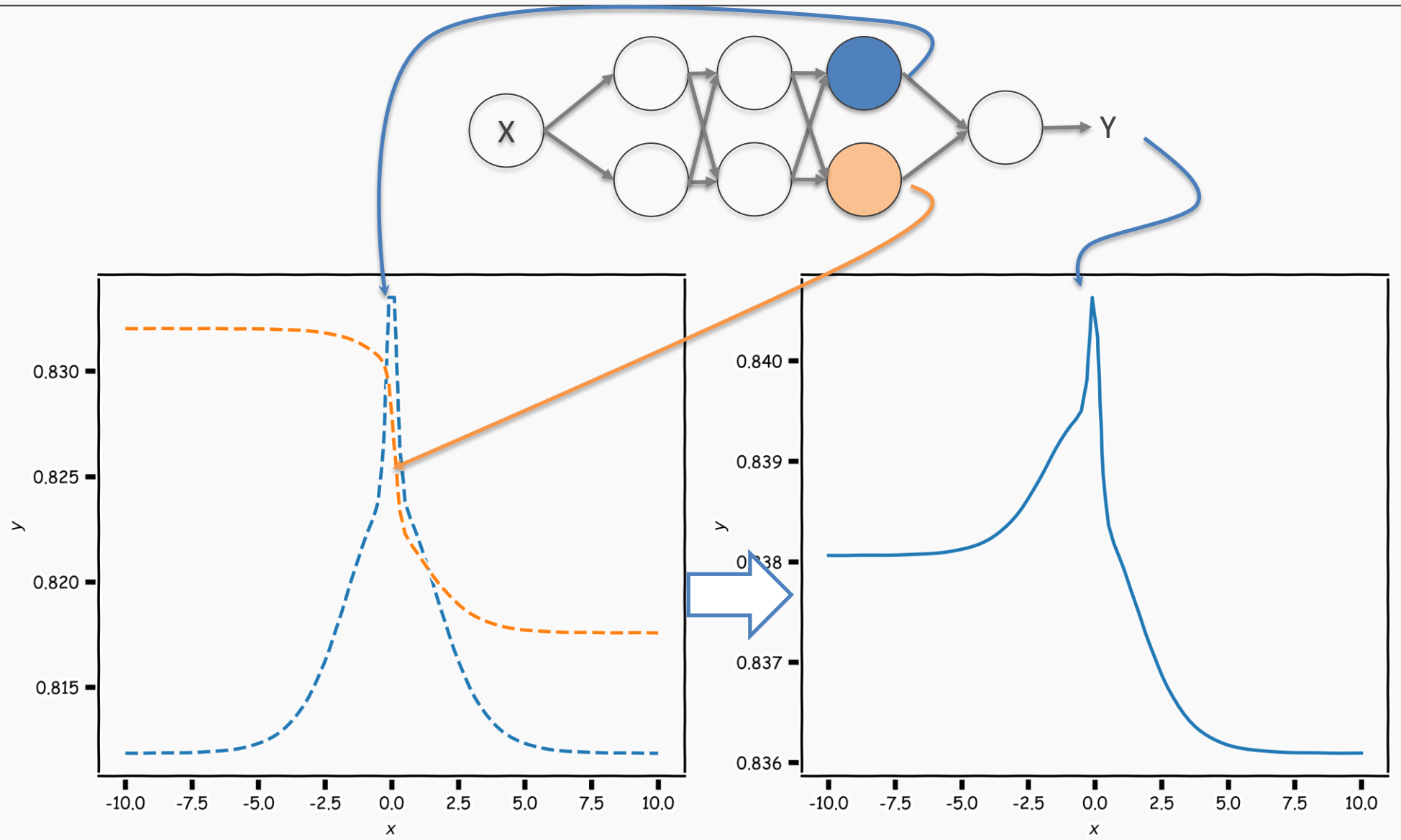
Different weights change the shape and position



Neural networks can model *any* reasonable function



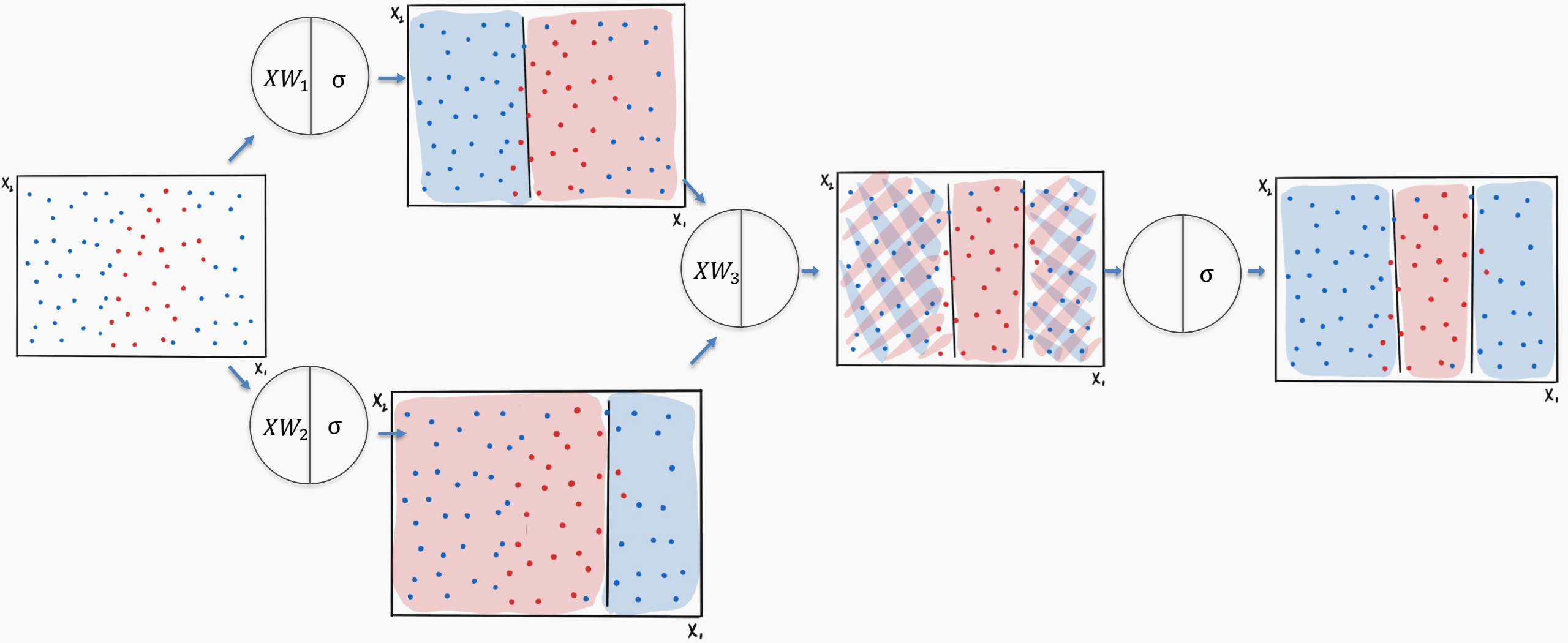
Neural networks can model *any* reasonable function



Neural networks can model *any* reasonable function

So, adding neurons and layers allows us to model increasingly complex functions and that is what a Neural Network is.

For 2-D input the same idea applies



Summary

So far:

- A single neuron can be a logistic regression or linear unit. We will soon see other choices of activation functions.
- A neural network is a combination of logistic regression (or other types) units.
- A neural network can approximate non-linear functions either for regression or classification.

Summary

Next:

- What kind of activations?
- How many neurons?
- How many layers?
- How to construct the output unit?
- What loss functions are appropriate?

Don’t worry it’s not as difficult as it sounds

Summary

Following two lectures on NN:

- How do we estimate the weights and biases?
- How to regularize Neural Networks?