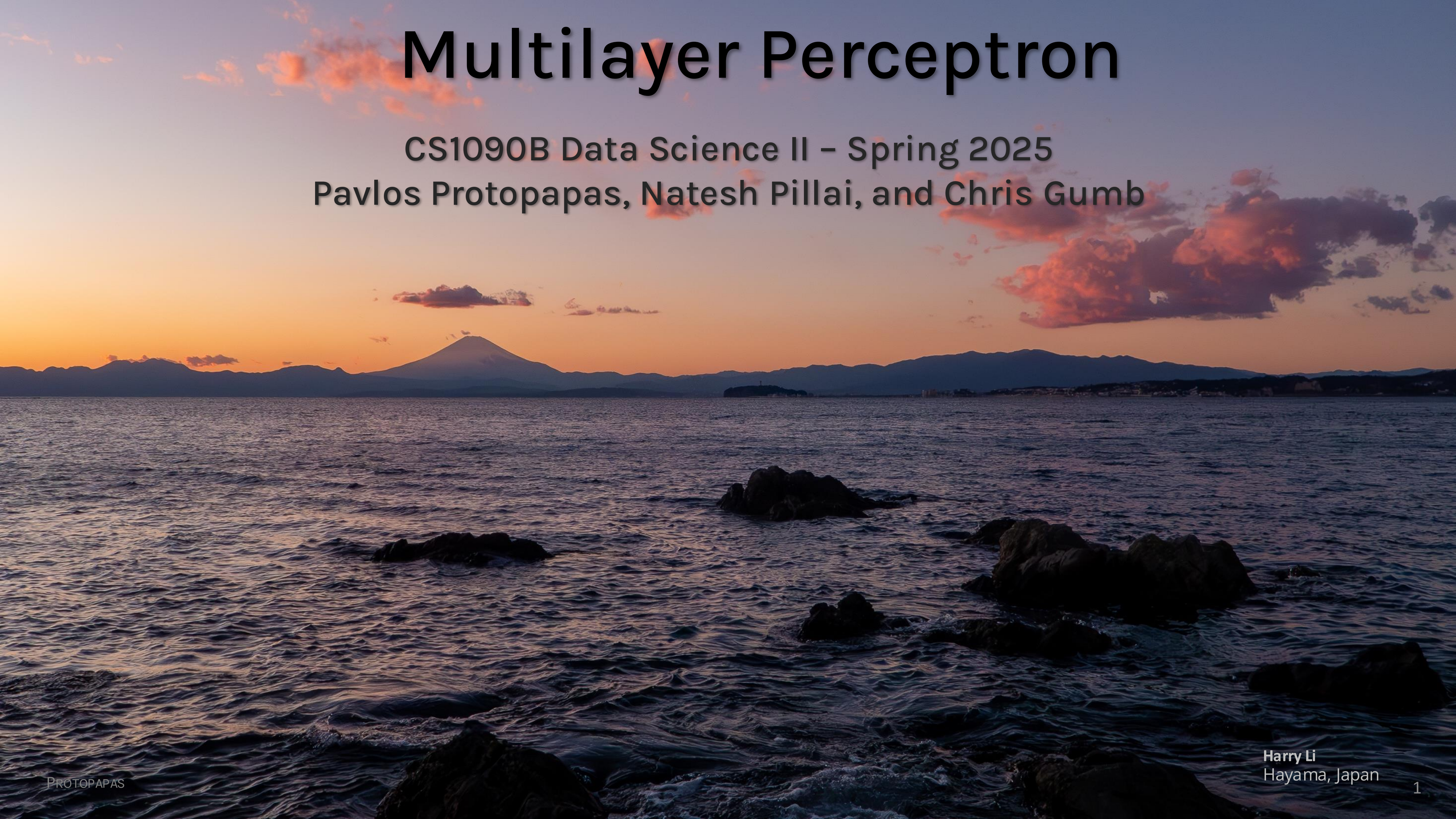
Multilayer Perceptron

CS1090B Data Science II – Spring 2025

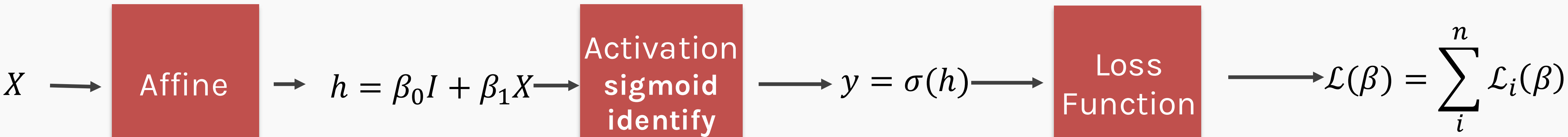
Pavlos Protopapas, Natesh Pillai, and Chris Gumb

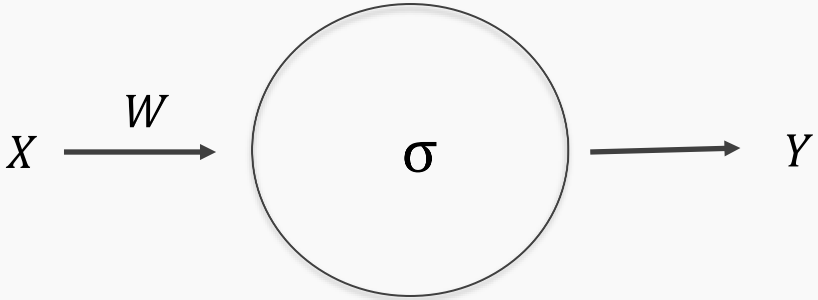


Harry Li
Hayama, Japan

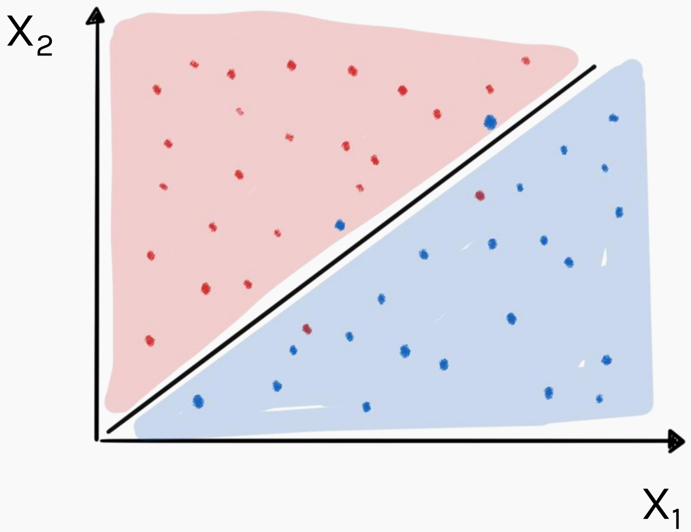
A single neuron

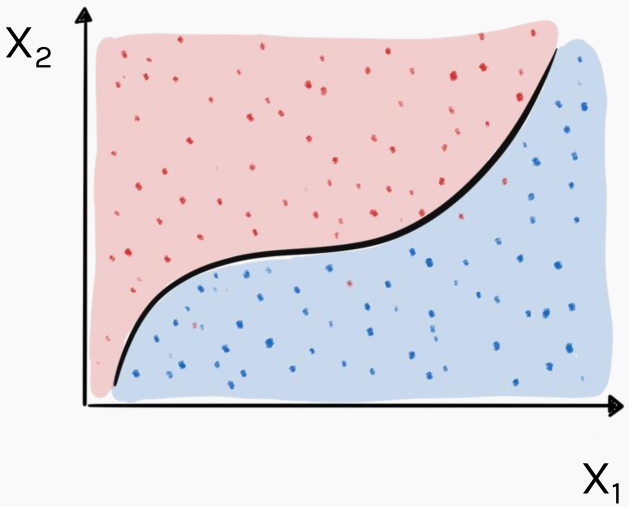
Up to this point we just re-branded logistic regression or linear regression to look like a neuron.

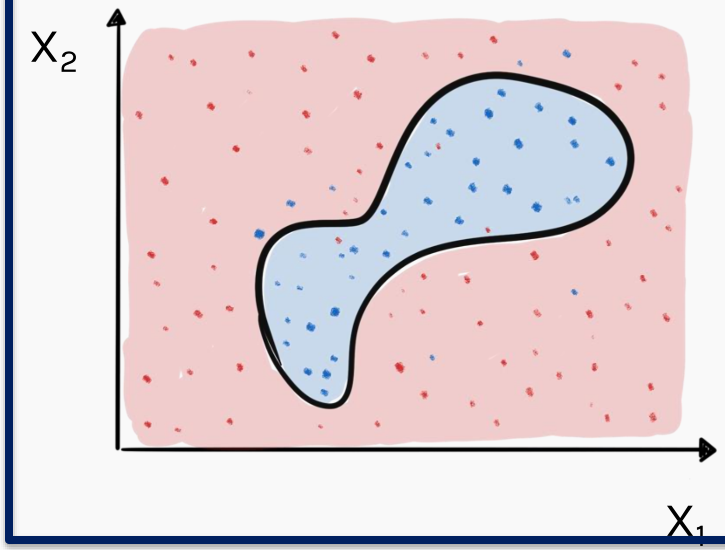


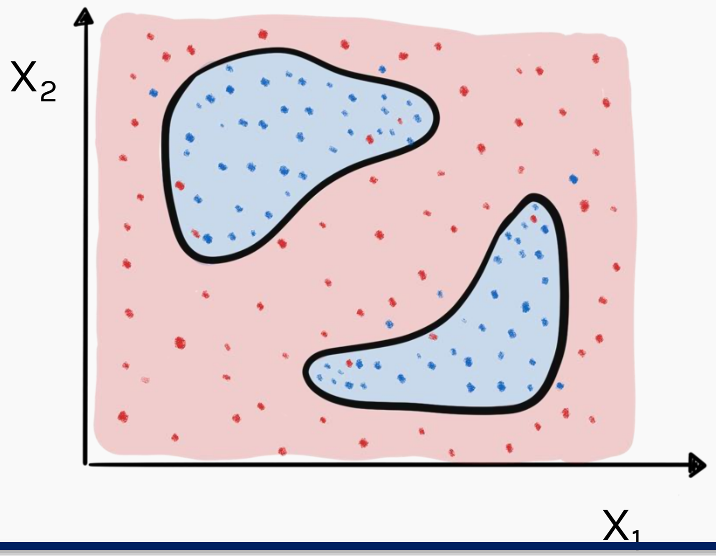


So, what's the big deal about Neural Networks?



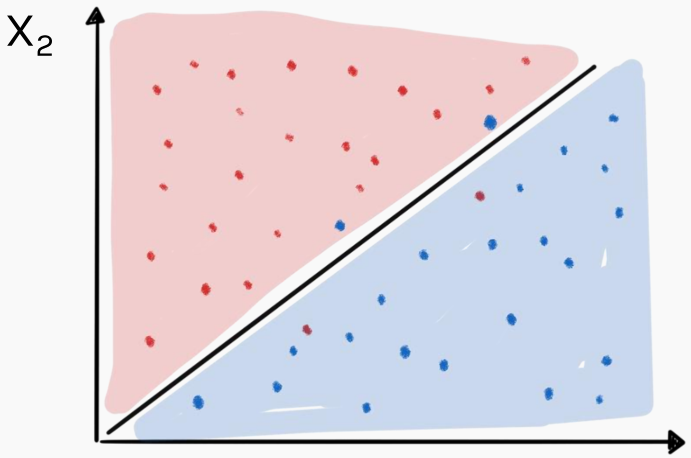


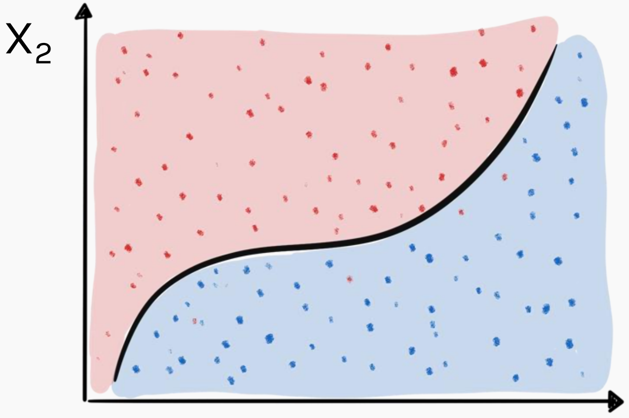




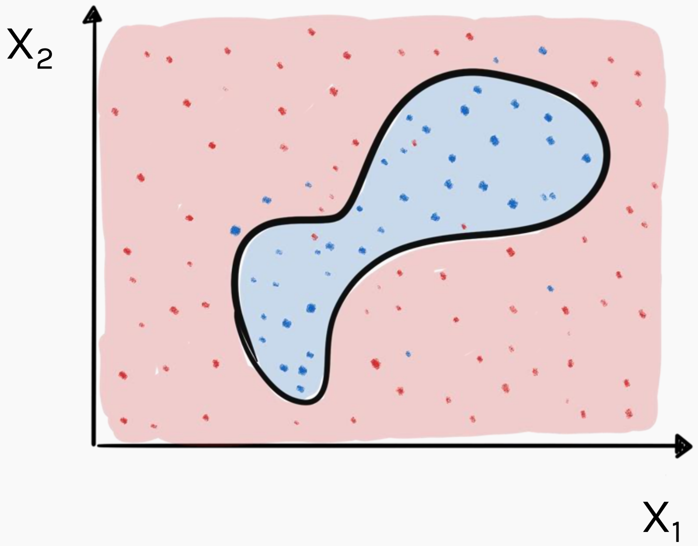
Very hard to express boundaries like these with logistic regression

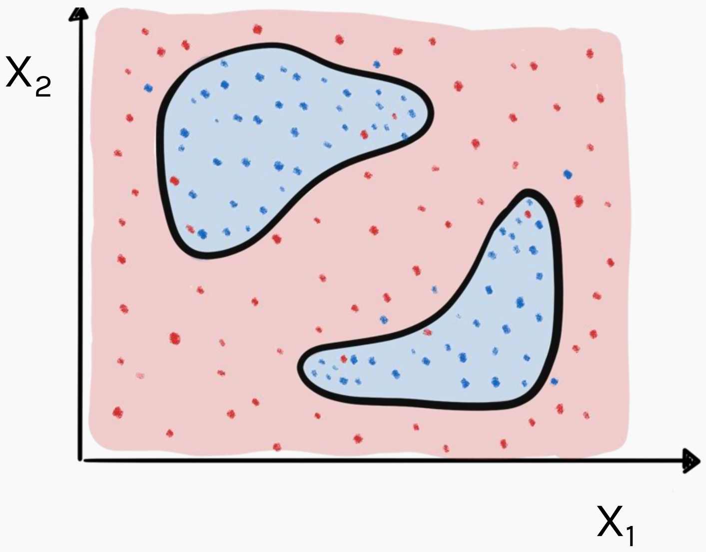
So, what's the big deal about Neural Networks?



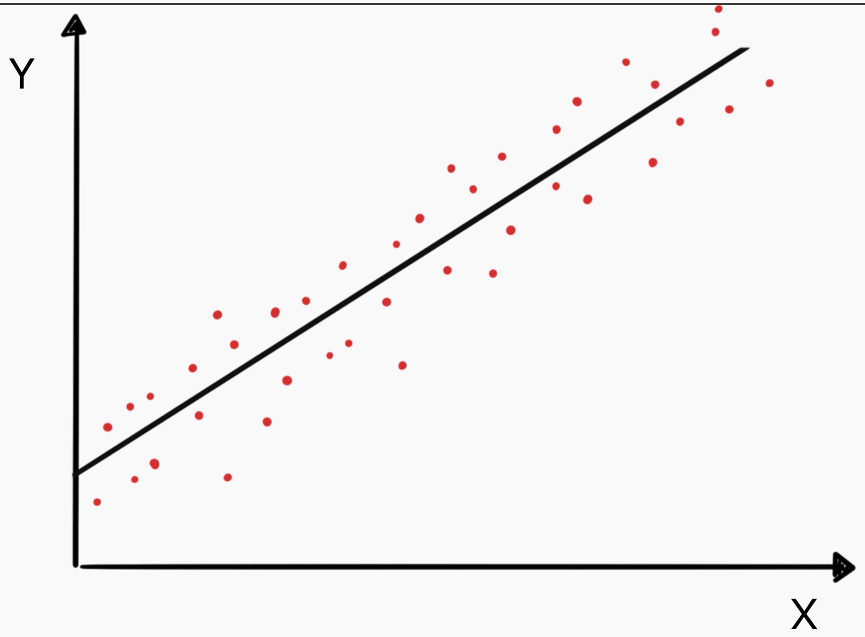


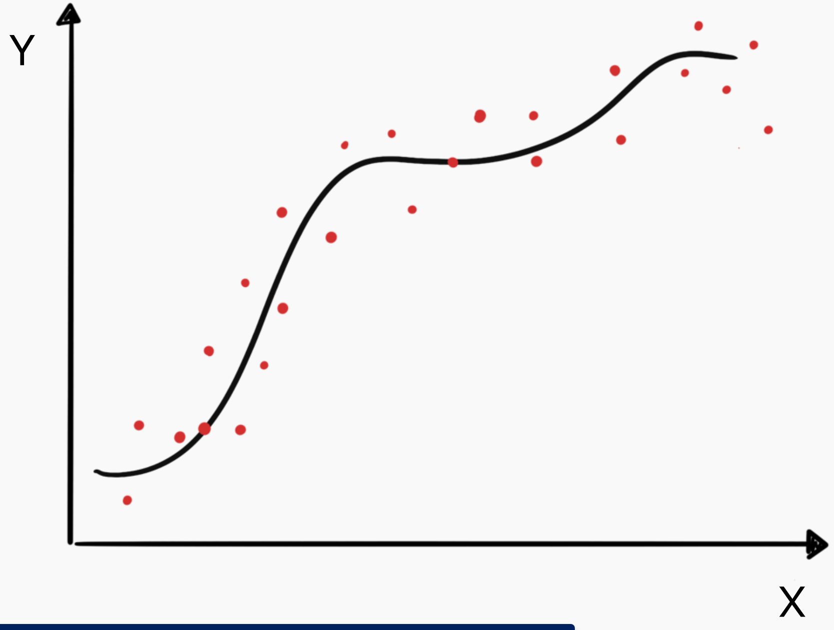
No Linear Boundaries!

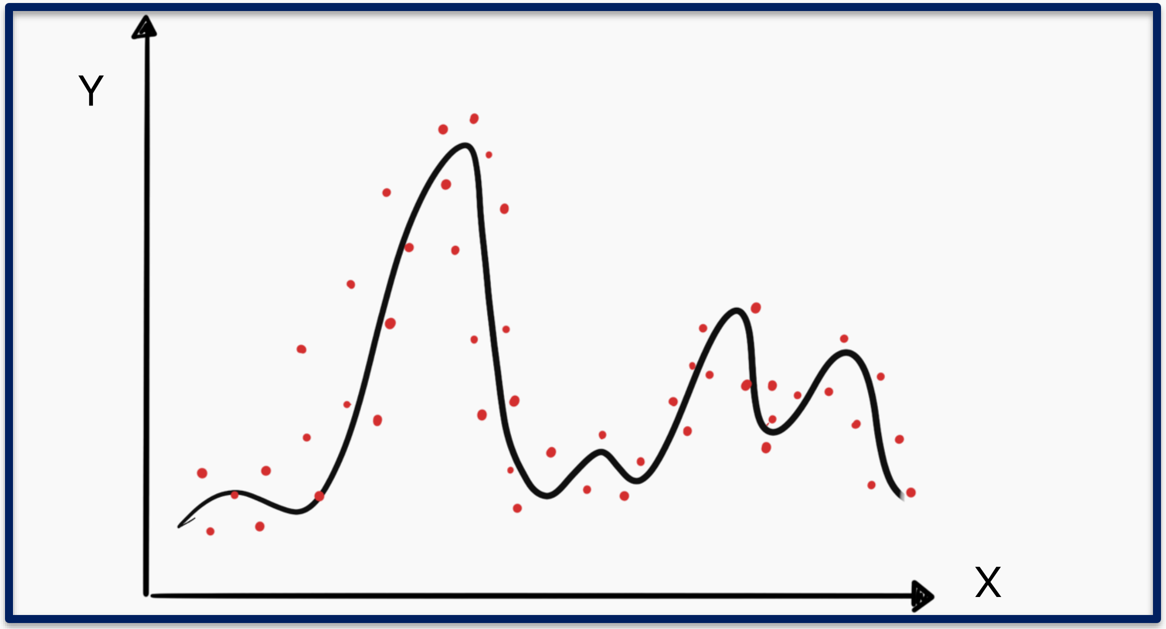




So, what's the big deal about Neural Networks?

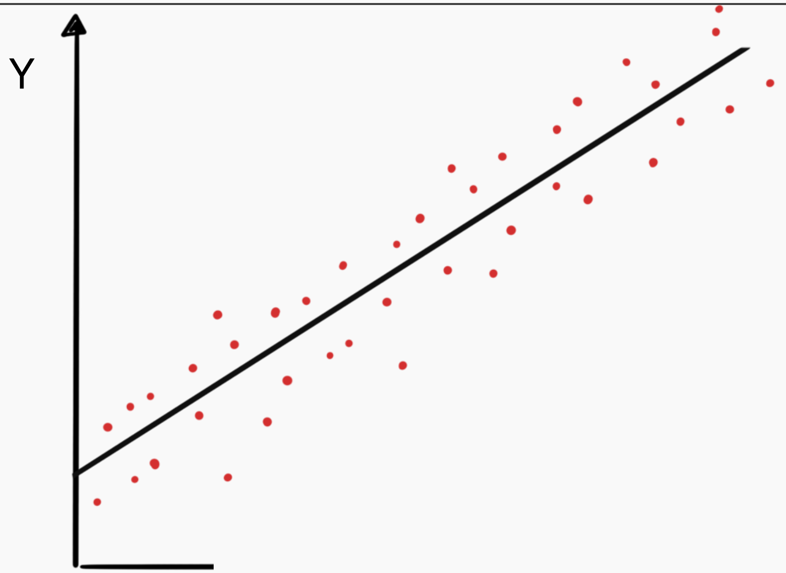


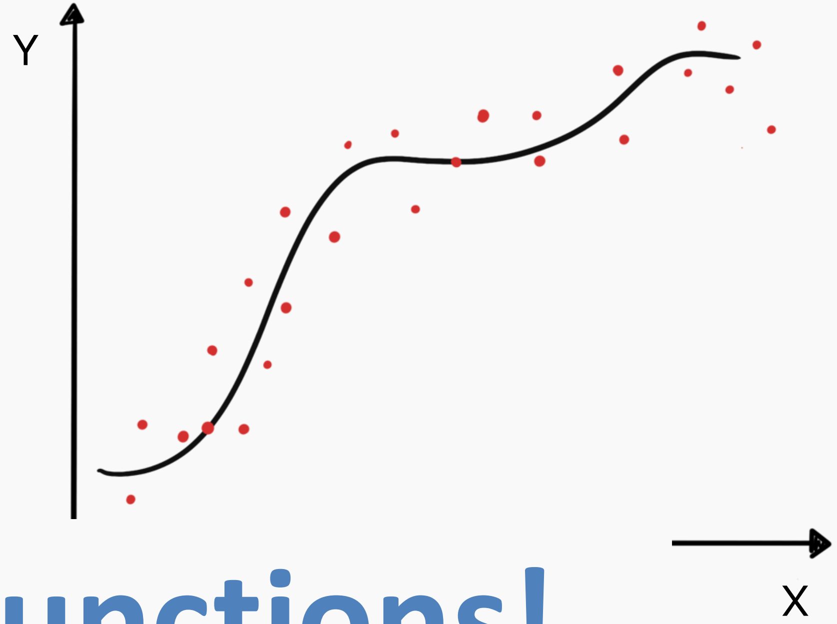




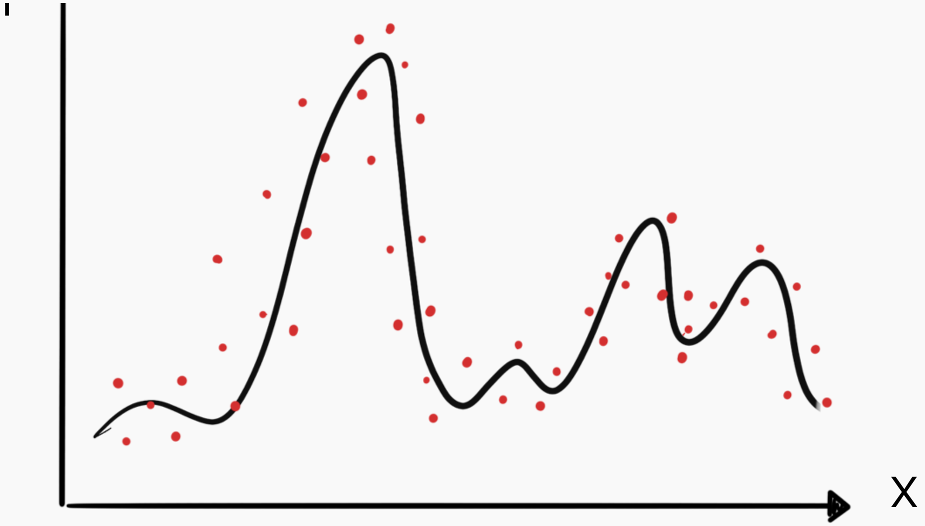
Very hard to express functions like these with polynomial regression

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





No Linear Functions!



Outline

- Introduction
- Review of basic concepts
- Linear to Logistic Regression
- Logistic Regression to ANN
- Perceptron - Single neuron network
- **Multi-Layer Perceptron (MLP)**

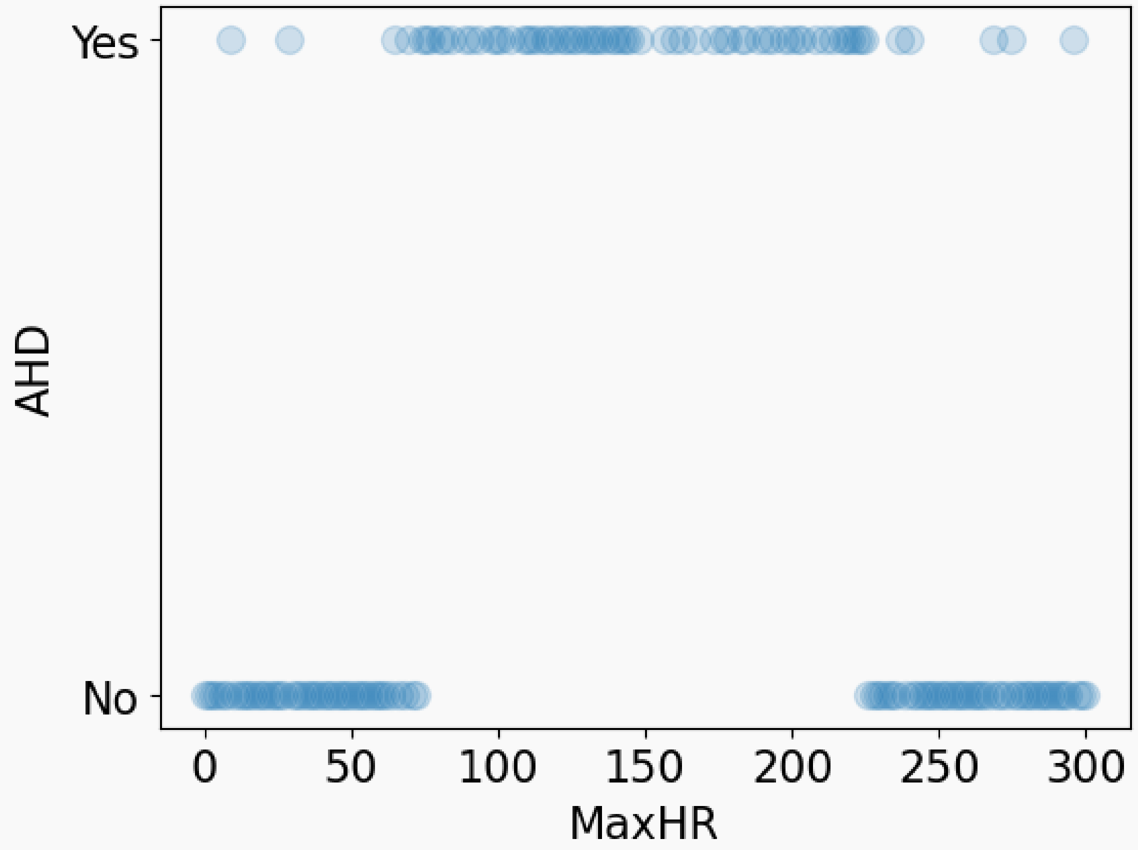
Example Using Heart Data

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

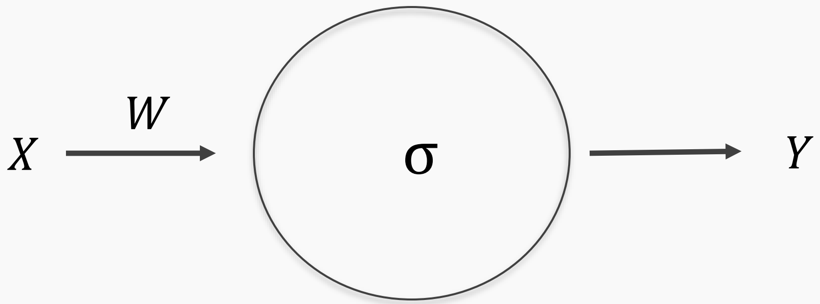
Age	Sex	ChestPain	RestBP	Chol	MaxHR	ExAng	Thal	AHD
63	1	typical	145	233	150	0	fixed	No
67	1	asymptomatic	160	286	108	1	normal	Yes
67	1	asymptomatic	120	229	129	1	reversable	Yes
37	1	nonanginal	130	250	187	0	normal	No

Example Using Heart Data

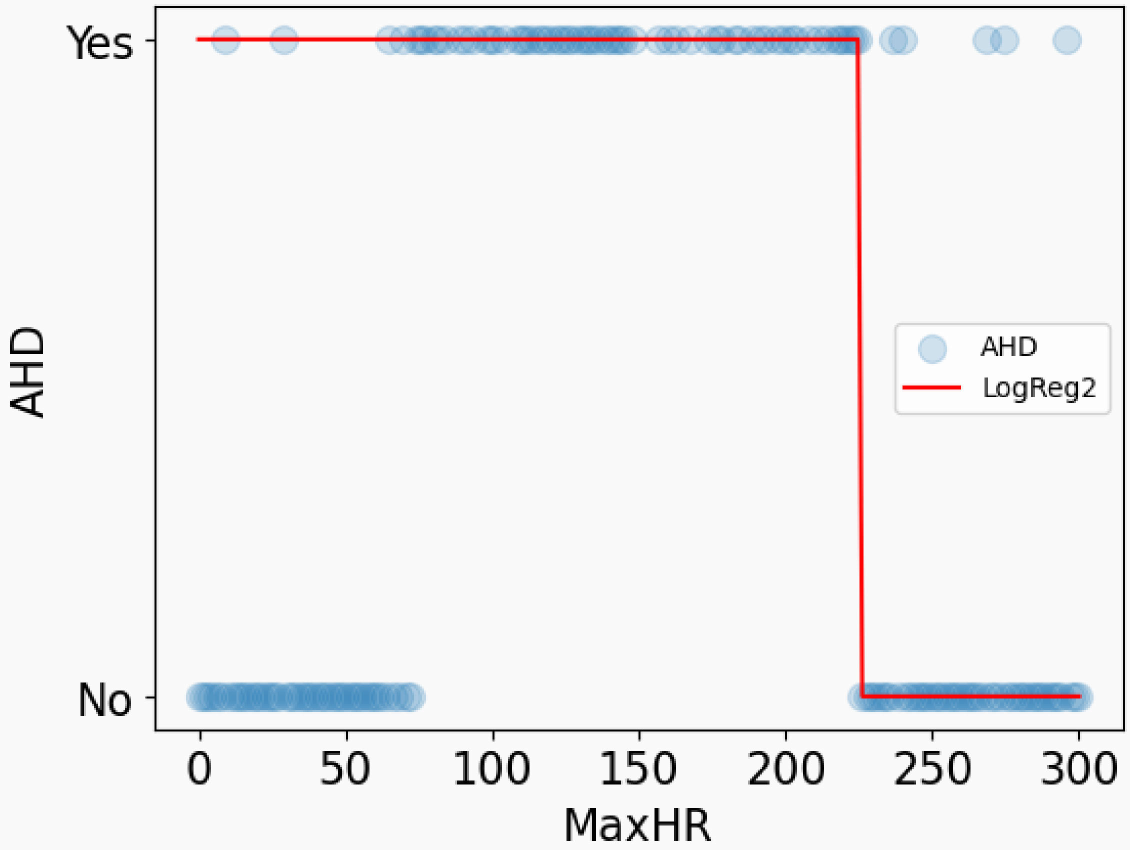
Slightly modified data to illustrate concepts.



Example Using Heart Data



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



Game Time

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

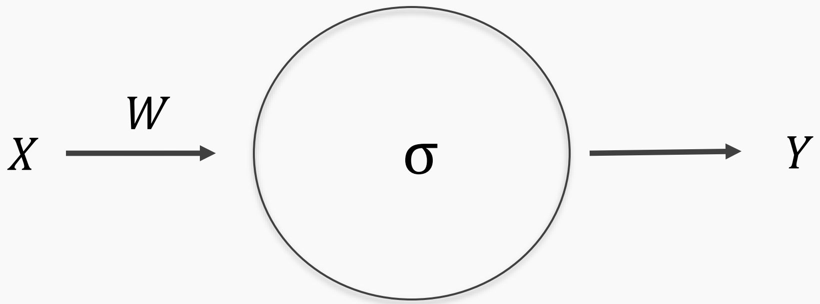
A. The curve will shift upwards

B. The curve will shift downwards

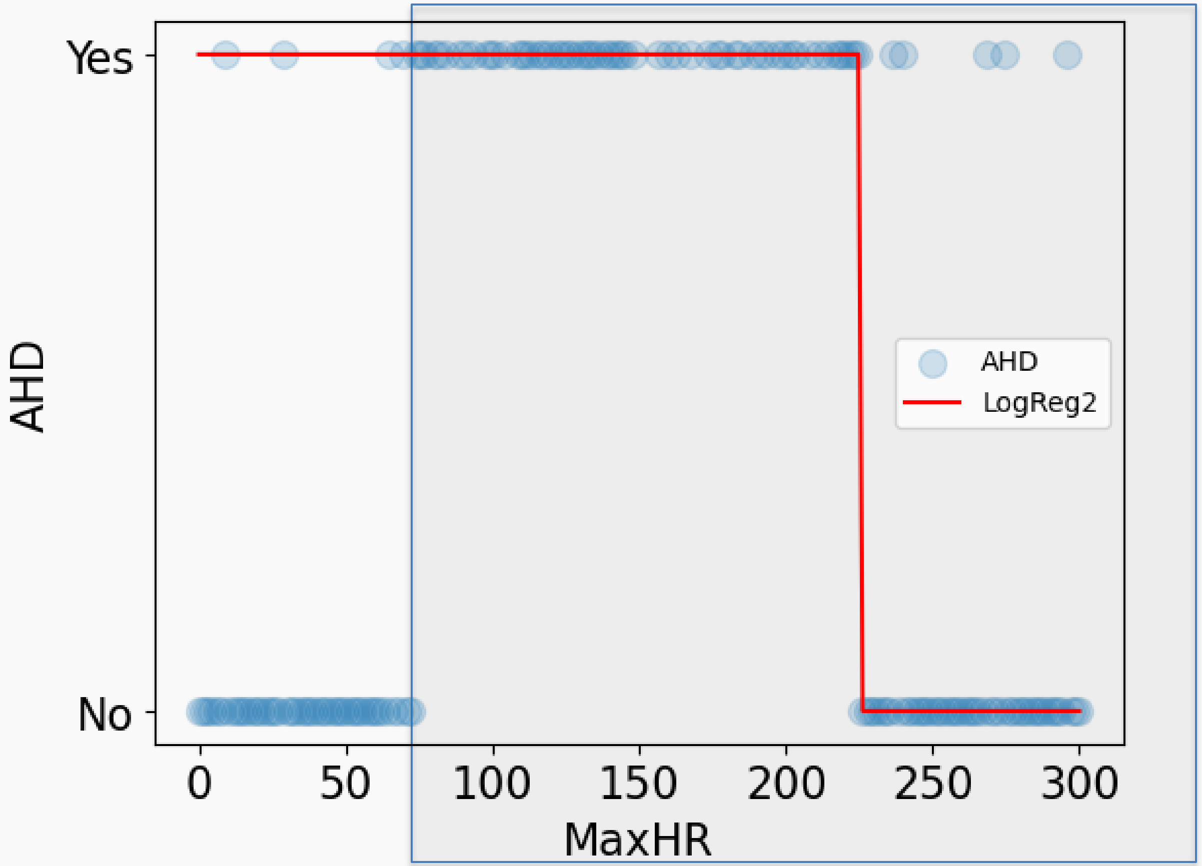
C. The curve will shift to the left

D. The curve's slope will become steeper

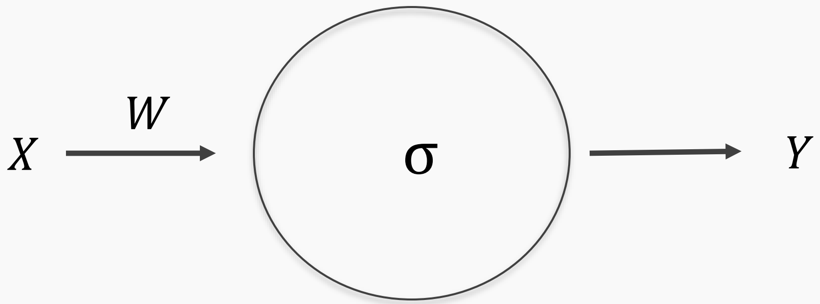
Example Using Heart Data



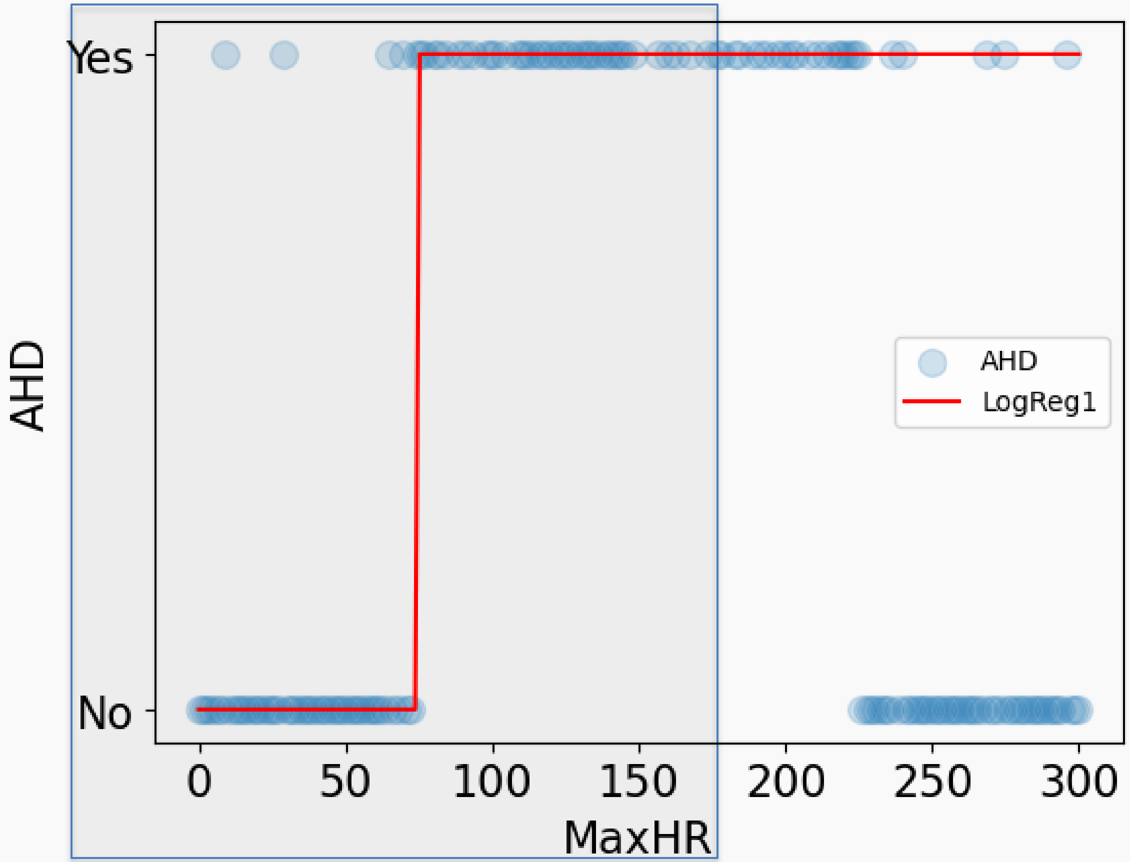
Choose W such as **right** part of data is fitted well



Example Using Heart Data

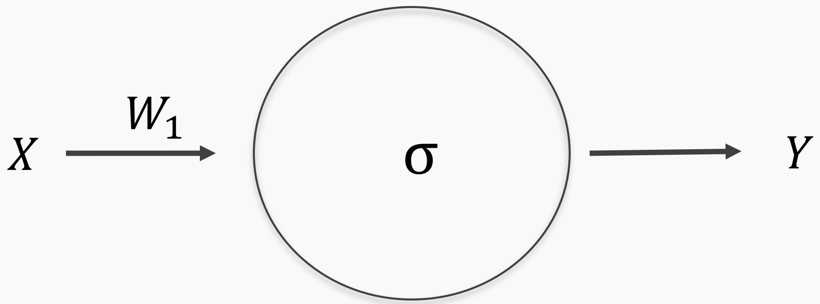


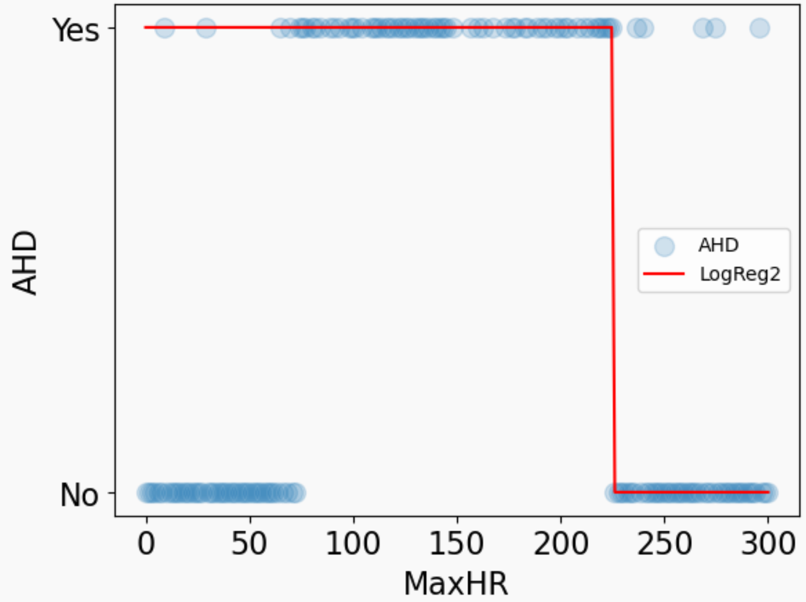
Choose W such as **left** part of data is fitted well

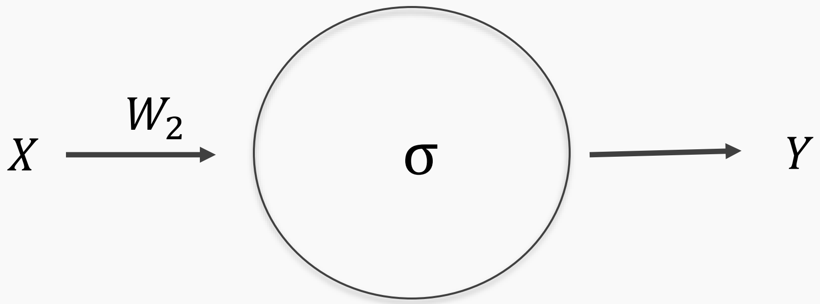


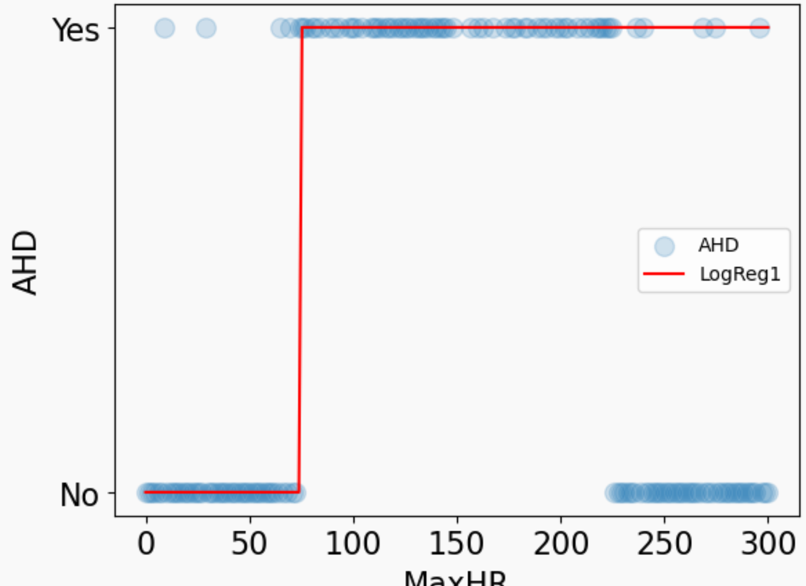
Example Using Heart Data

Two regions, two nodes

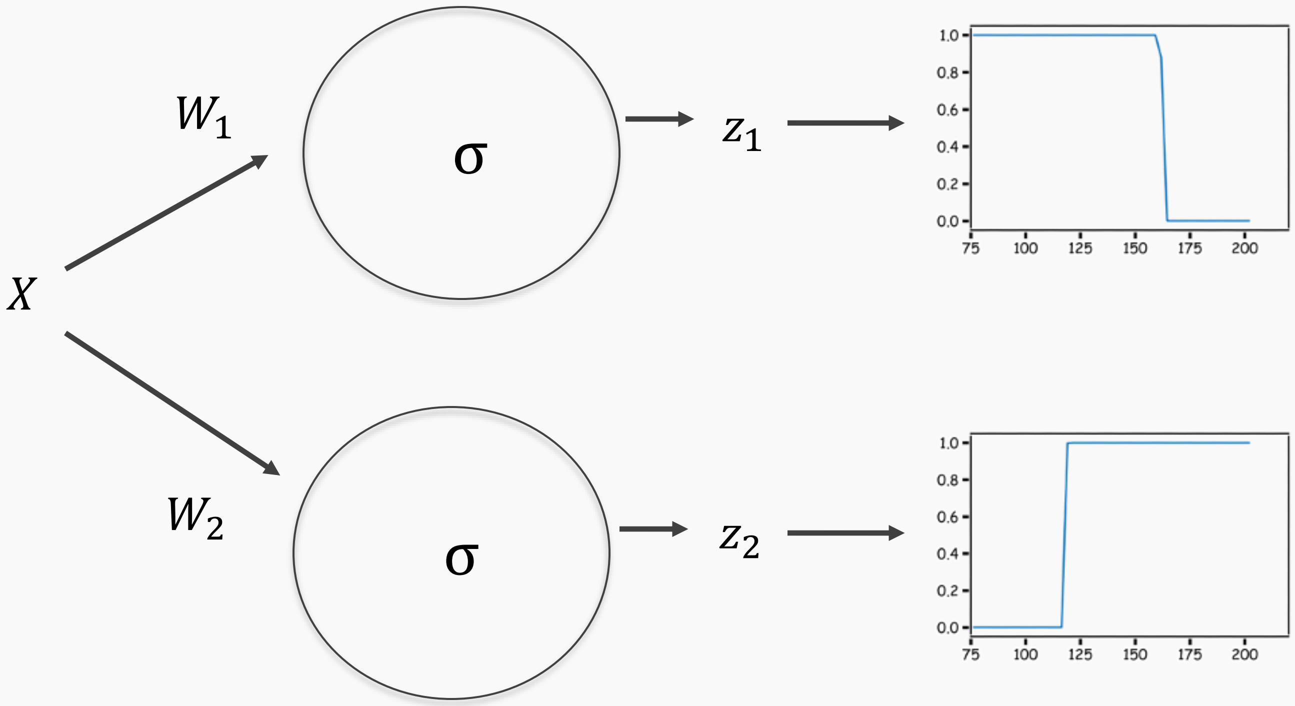








Combining Neurons



Game Time

How shall we combine these results?

A. Multiply z_1 and z_2

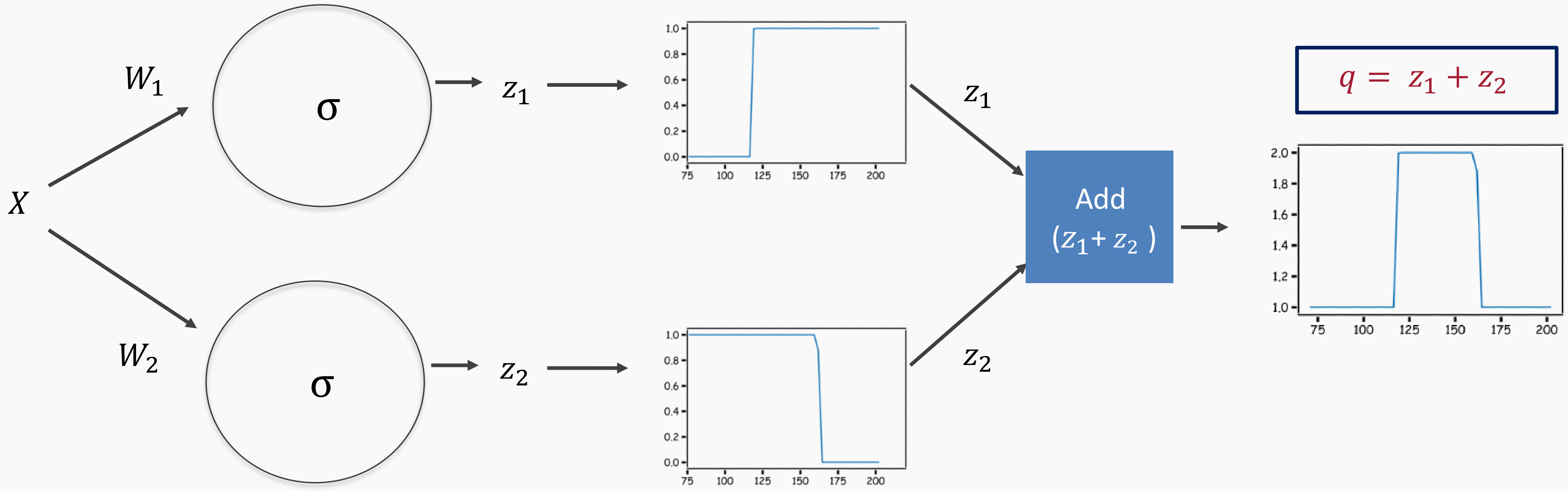
B. Add z_1 and z_2

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

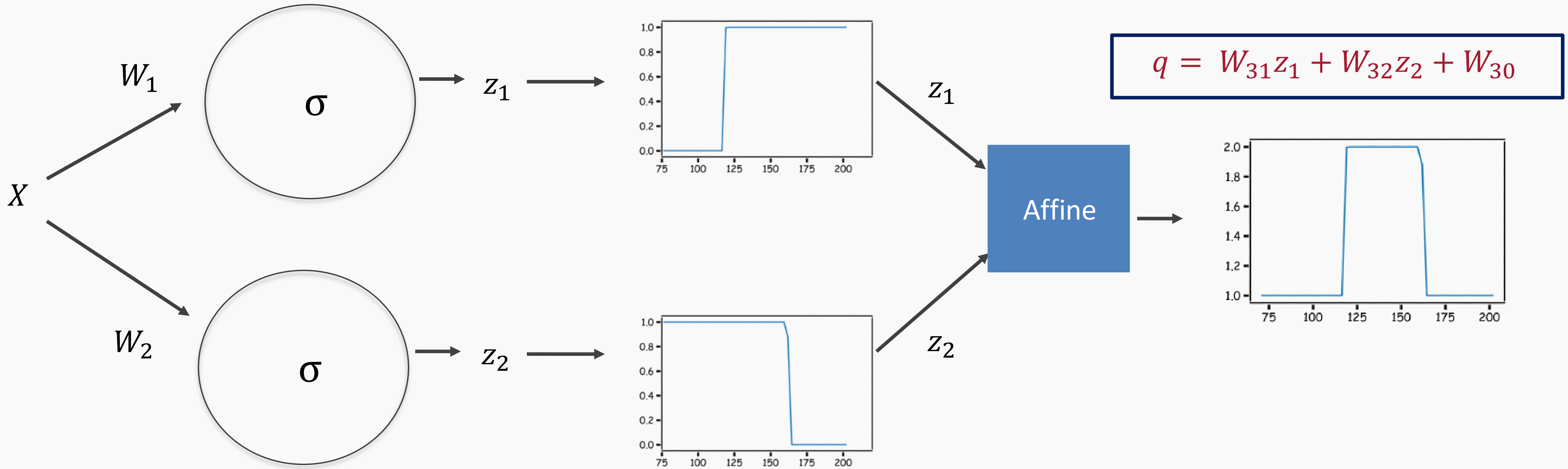
D. Convolve z_1 with z_2

 17

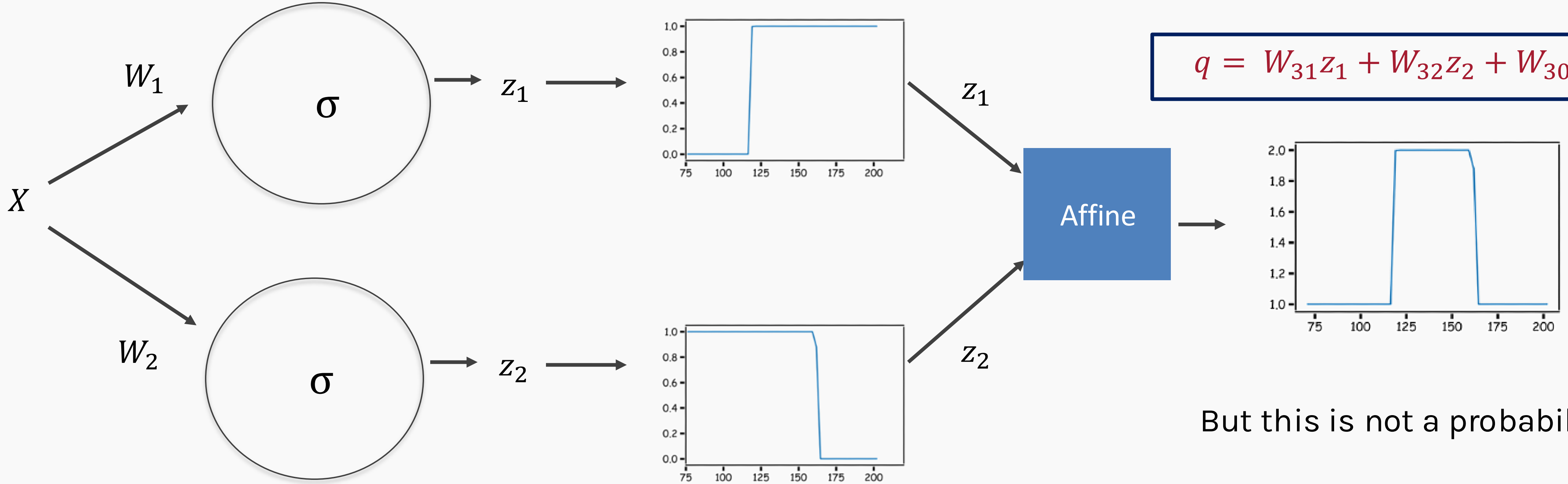
Combining Neurons



Combining Neurons



Combining Neurons



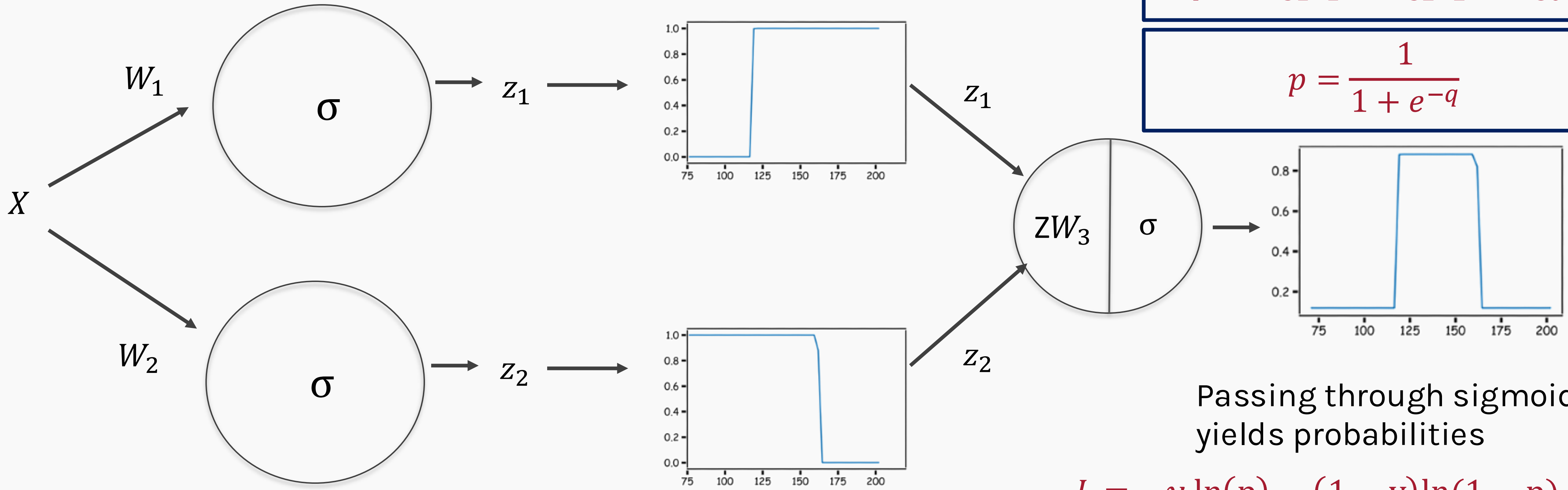
$$q = W_{31}z_1 + W_{32}z_2 + W_{30}$$

But this is not a probability

Combining Neurons

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

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



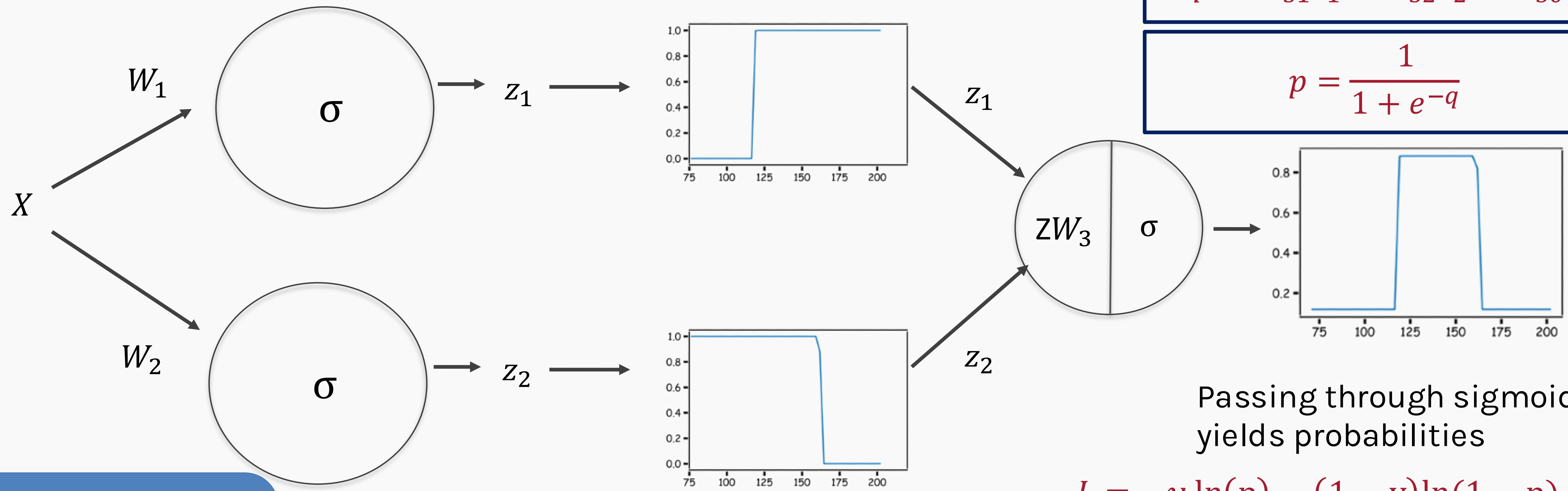
Passing through sigmoid yields probabilities

$$L = -y\ln(\mathrm{p}) - (1-\mathrm{y})\ln(1-\mathrm{p})$$

Combining Neurons

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

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



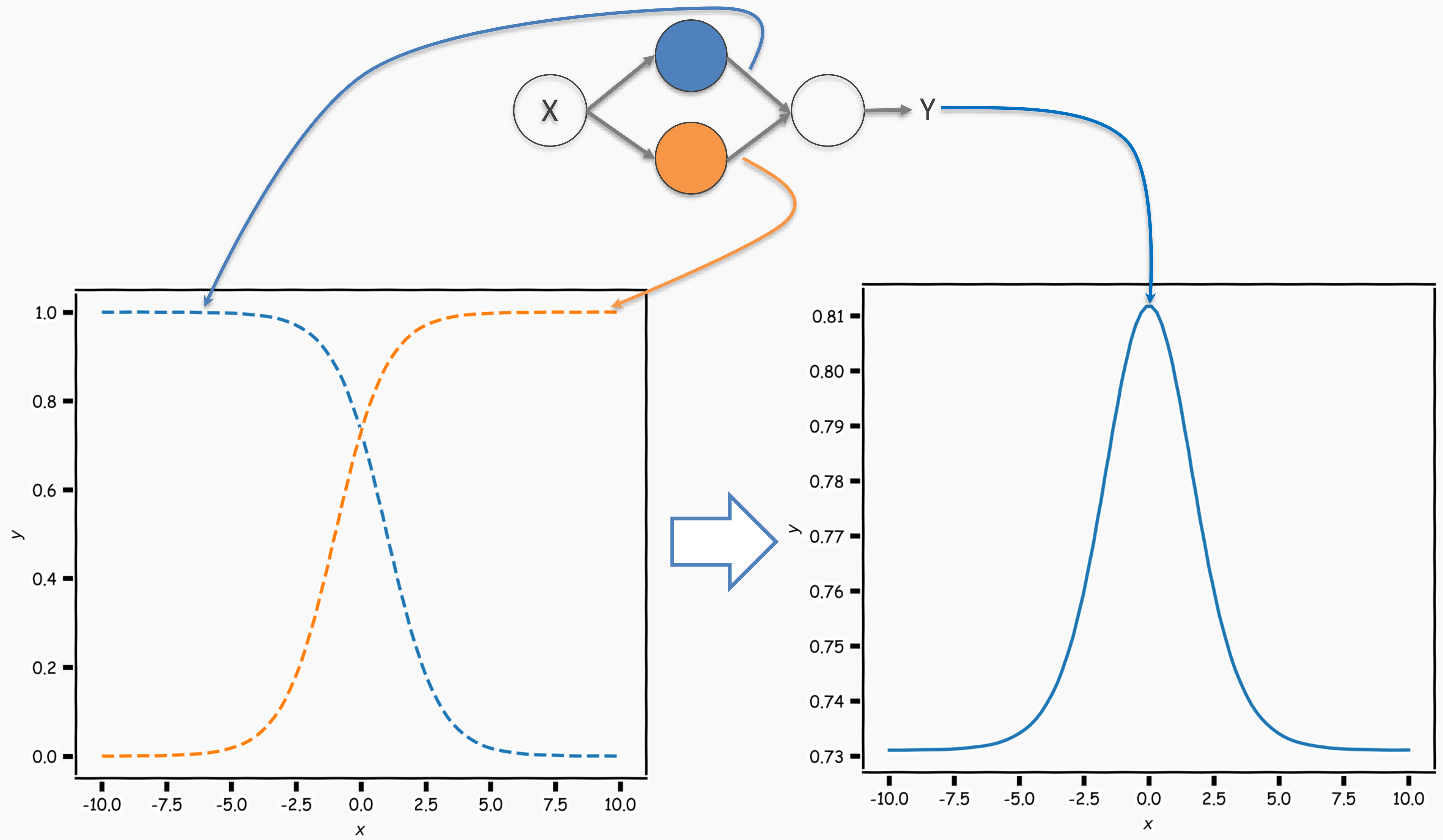
Passing through sigmoid yields probabilities

$$L = -y \ln(\mathrm{p}) - (1 - \mathrm{y}) \ln(1 - \mathrm{p})$$

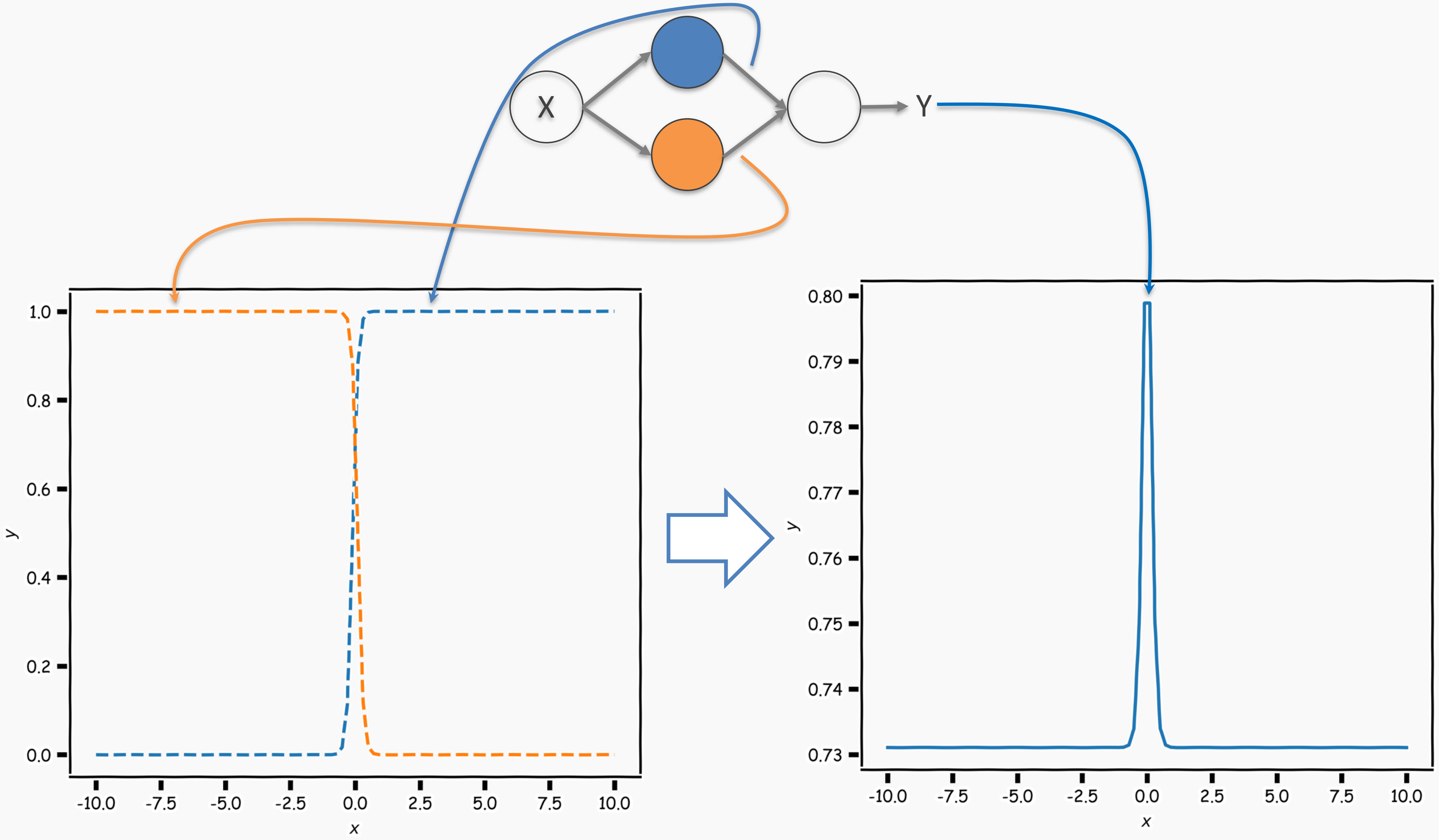
We will talk about this in the coming lectures

Need to learn W_1 ,W_2 and W_3

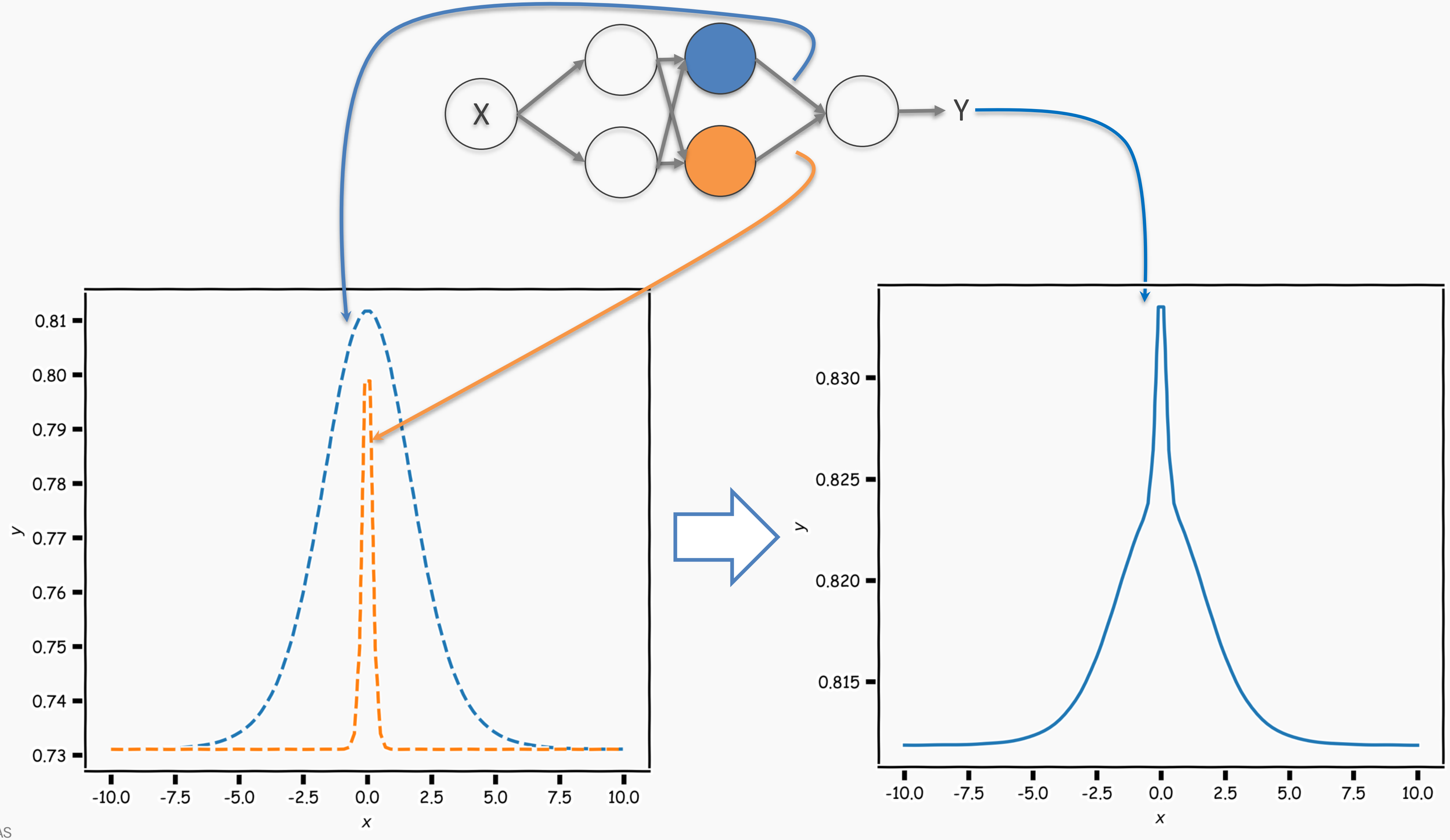
Combining neurons allows us to model interesting functions



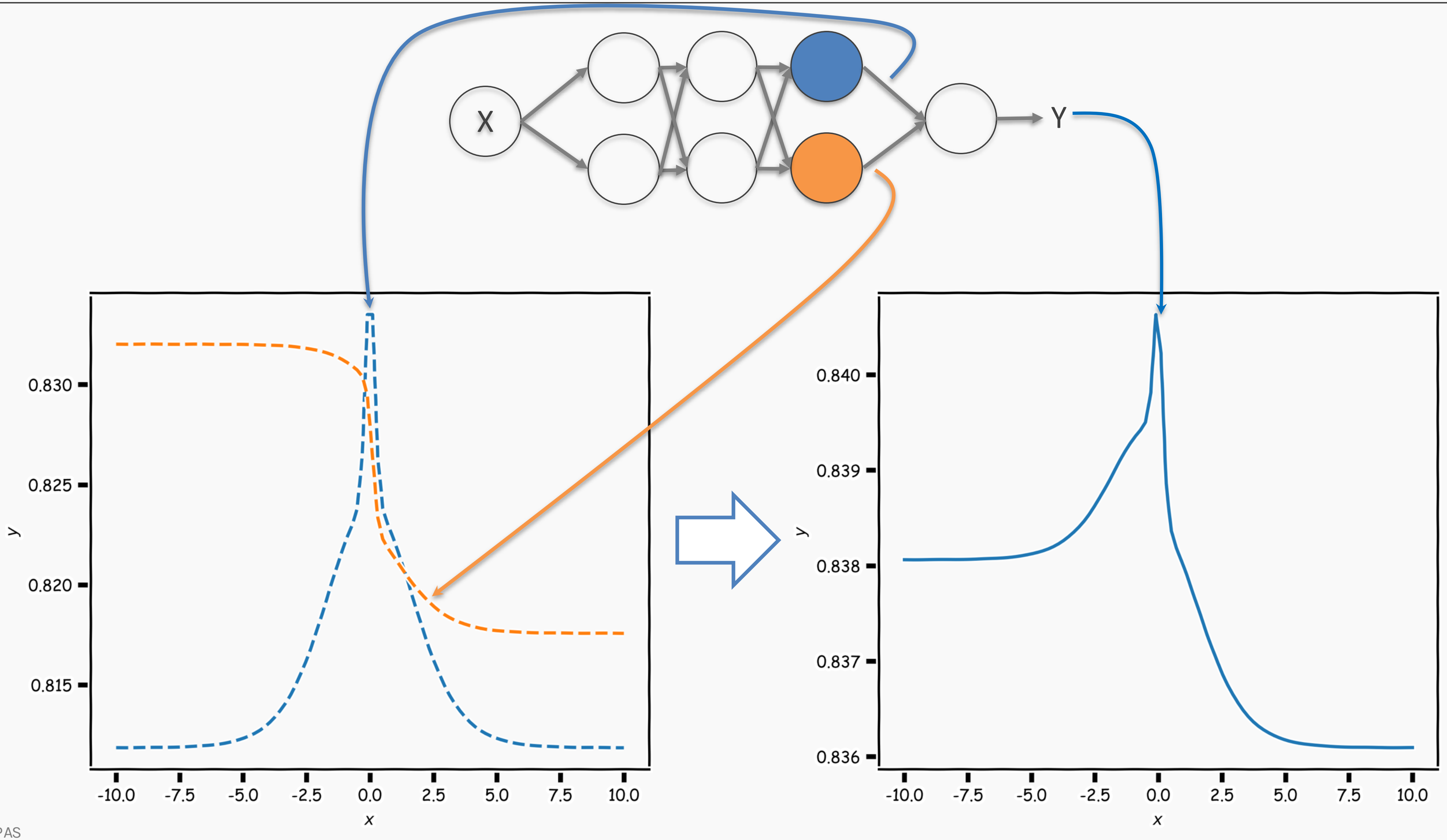
Combining neurons allows us to model interesting functions



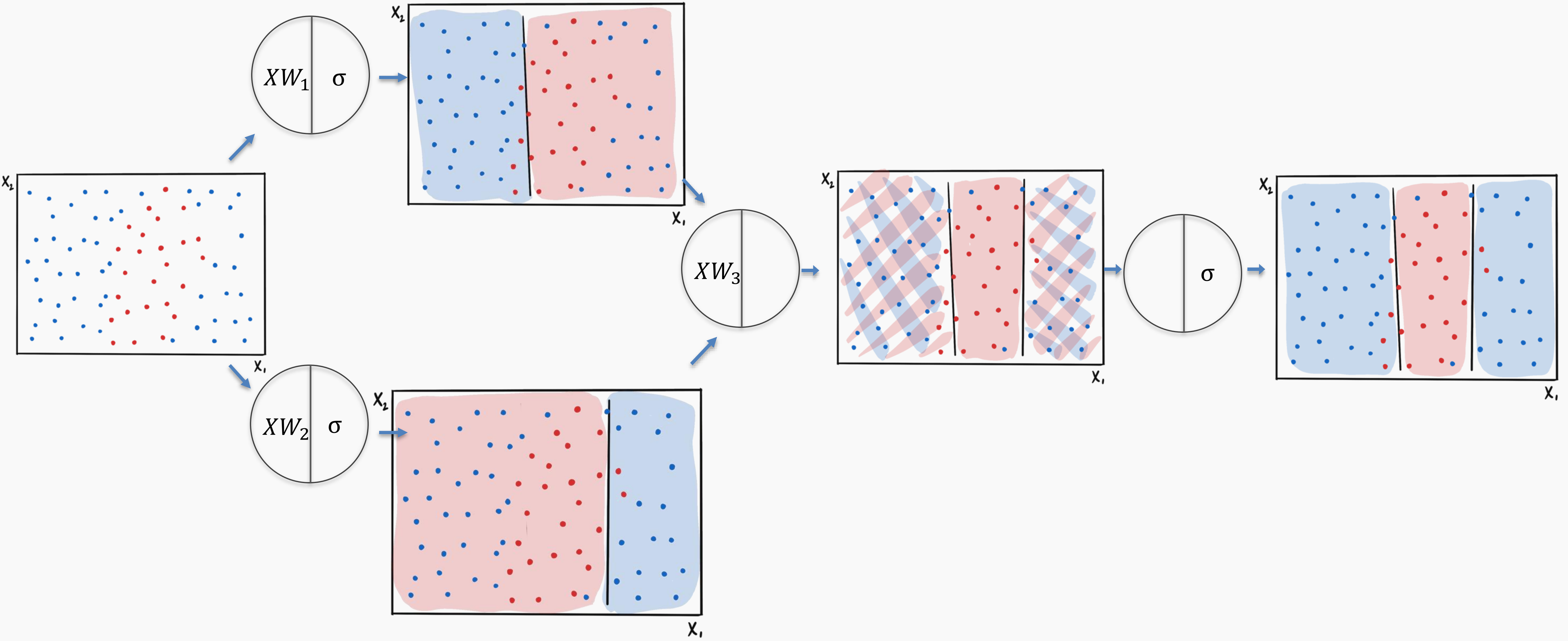
Combining neurons allows us to model interesting functions



Combining neurons allows us to model interesting functions



For 2-D input the same idea applies



Summary

So far:

- A single neuron is simply an affine transformation followed by an activation function.
- A single neuron can function either as a logistic regression unit or linear unit. We will soon see other choices of activation functions.
- A neural network is a combination of neurons 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 and what loss functions are appropriate