Introduction to Artificial Intelligence
Single-layer Neural Networks

Martha Lewis

Department of Engineering Mathematics
University of Bristol
martha.lewis@bristol.ac.uk

Single Layer Neural Networks

Q: When was the first neural network architecture proposed?

Single Layer Neural Networks

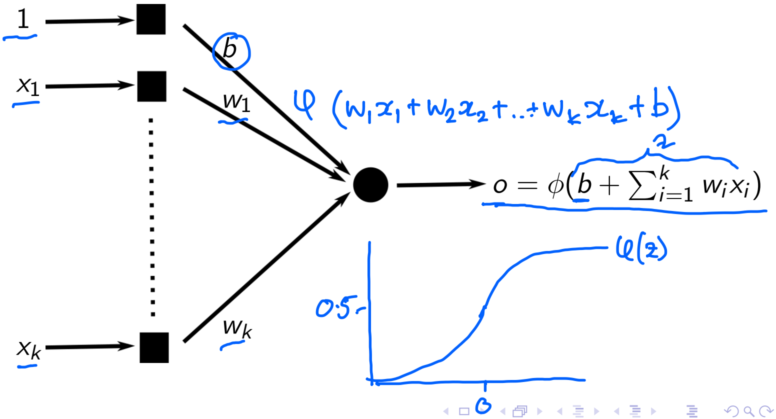
Q: When was the first neural network architecture proposed?

A: McCulloch and Pitts first proposed a model artificial neural networks in 1943.
Rosenblatt then proposed the Perceptron in 1958.
For a short history, have a look at `https://www.ibm.com/uk-en/cloud/learn/neural-networks#toc-history-of-rIfu5uF2`

Perceptron Model

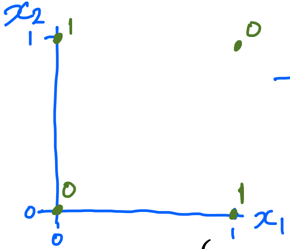
$(\vec{x}, y)$

- Suppose we have a data set of labelled examples $(x_1, \ldots, x_k, y)$ where $x_i \in \mathbb{R}$ for $i = 1, \ldots, k$ and also $y \in \mathbb{R}$.
- In these examples labels and attributes are related by an *unknown* function $f : \mathbb{R}^k \to \mathbb{R}$ such that $y = f(x_1, \ldots, x_k)$.

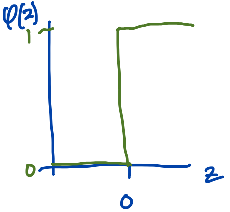


Limitations of Single Layer NNs

Consider the XOR function $f : \{0,1\}^2 \to \{0,1\}$ defined by the following table:



x_1	x_2	$f(x_1, x_2)$
1	1	0
1	0	1
0	1	1
0	0	0



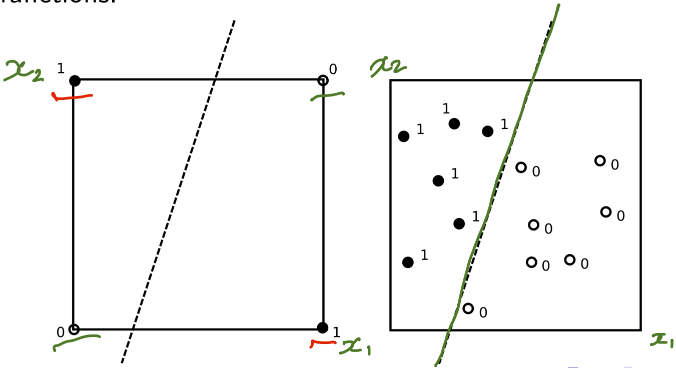
Let $\phi(z) = \begin{cases} 0 : z \leq 0 \\ 1 : z > 0 \end{cases}$ then for a single-layer neural network we

$x_1 = 1,\ x_2 = 1$

require $\phi(w_1 x_1 + w_2 x_2) = f(x_1, x_2)$.
This means we need that: $\phi(w_1 + w_2) = 0$, $\phi(w_1) = 1$, $\phi(w_2) = 1$ and $\phi(0) = 0$.
But then, $w_1 + w_2 \leq 0$ and $w_1 > 0$ and $w_2 > 0$. These constraints are inconsistent!

Linearly Separable

- Single-layer NNs can only solve linearly separable problems.
- A problem with n inputs is linearly separable if it is possible to find a n-dimensional hyperplane which geometrically separates the sets.
- This is a serious restriction e.g. there are 2^{2^n} n-dimensional Boolean functions and only of order 2^{n^2} separable Boolean functions.



Linearly separable problems - in 2D

- The equation for a line is...

Linearly separable problems - in 2D

- The equation for a line is...
- $y = mx + c$

Linearly separable problems - in 2D

- The equation for a line is...
- $y = mx + c$
- Suppose we have a single layer neural network on a 2D space with a step function as non-linearity (not sigmoid)

$$\boxed{\phi([x_1, x_2])} = \begin{cases} 1 & \text{if } w_1x_1 + w_2x_2 + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Linearly separable problems - in 2D

- The equation for a line is...
- $y = mx + c$
- Suppose we have a single layer neural network on a 2D space with a step function as non-linearity (not sigmoid)

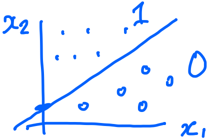
$$\phi([x_1, x_2]) = \begin{cases} 1 & \text{if } w_1x_1 + w_2x_2 + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Then:

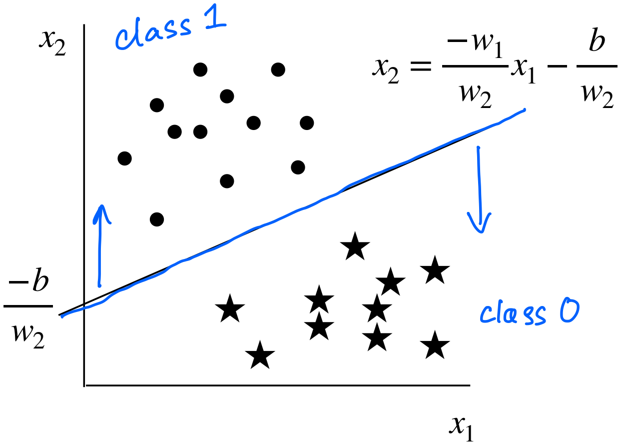
$$\phi([x_1, x_2]) = \underline{w_1x_1 + w_2x_2 + b} > 0 \iff x_2 + \frac{w_1}{w_2}x_1 + \frac{b}{w_2} > 0$$

divide by w_2

$$\iff x_2 > -\frac{w_1}{w_2}x_1 - \frac{b}{w_2}$$



Linearly separable problems - in 2D



Summary

- The first artificial neural networks used a single layer
- This kind of architecture was quickly shown to be inadequate since only linearly separable problems can be solved.
- In the 2D case we can calculate the line that the network models.
- Adding more layers allows a wider range of problems to be solved - in principle with one hidden layer any continuous function can be approximated