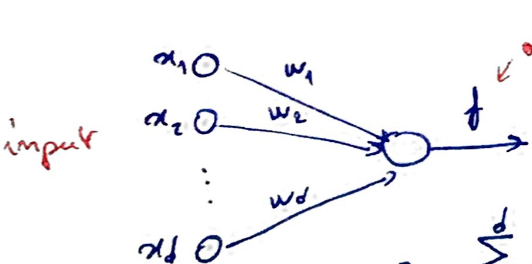
Unit 3 - Lecture 8: Introduction to Feedforward Neural Networks

Objectives:

- Recognize different layers in a Feedforward neural network and the number of units in each lay
- Write down common activation functions such as the hyperbolic tangent function tanH, and the rectified linear function (ReLu)
- Compute the output of a simple neural network possibly with Hidden layers given the weights and activation functions.
- Determine whether data after transformation by some layers is linearly separable, draw decision boundaries given by the weight vectors and use them to help understand the behavior of the network.

Neural Network Units:

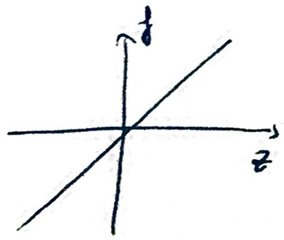
Neural networks are models in which the feature representation is learned jointly with the classifier to improve classification performance.



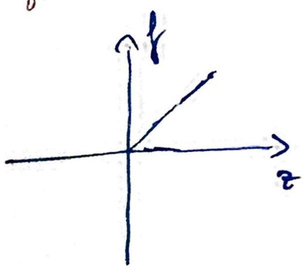
output $f(z)$ (non linear) called Relu.

$$z = \sum_{i=1}^{d} x_i \cdot w_i + w_0$$

number   weights

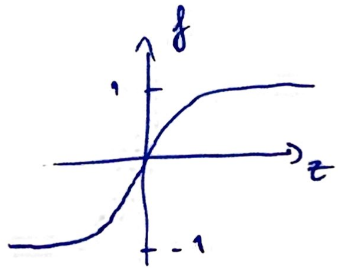


linear

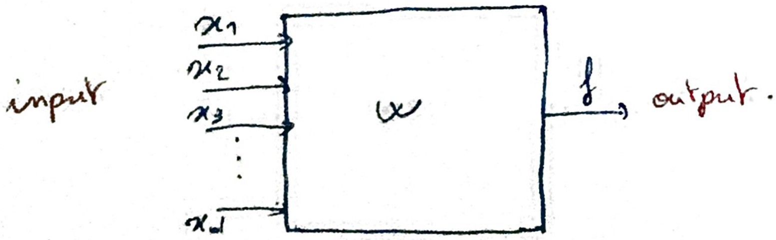


Relu

$f(z) = \max\{0, z\}$



$$\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

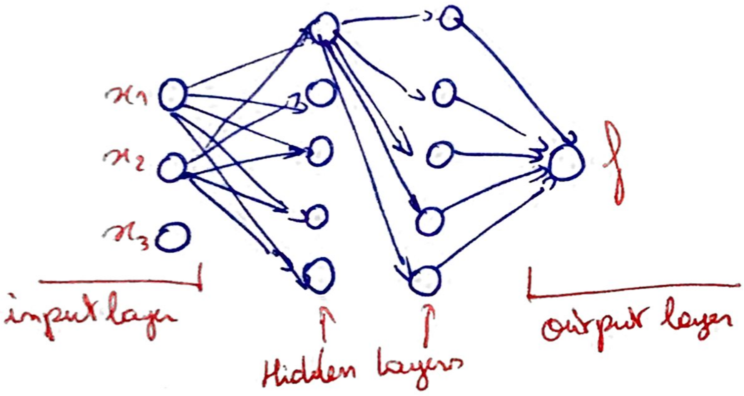


Introduction to Deep Neural Networks:

Why Deep learning?

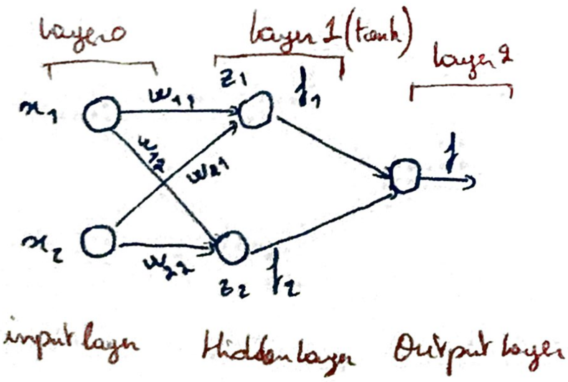
1. Lots of Data
2. Computational Resources (GPU)
3. Large models are easier to train
4. flexible neural "lego pieces"

A Deep (feed forward) neural Network refers to a neural network that contains not only the input and output layers, but also hidden layers in between. For example, below is a deep feedforward neural Network of 2 hidden layers, with each hidden layer consisting of 5 units:



Hidden layer Models:

One Hidden layer model:

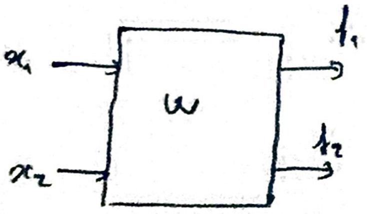


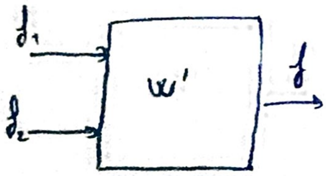
$$z_1 = \sum_{j=1}^{2} x_j w_{j1} + w_{01}$$

$$f_1 = f(z_1) = \tanh(z_1)$$

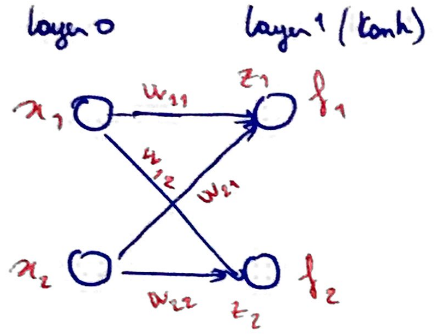
$$z_2 = \sum_{j=1}^{2} x_j w_{j2} + w_{02}$$

$$f_2 = f(z_2) = \tanh(z_2)$$

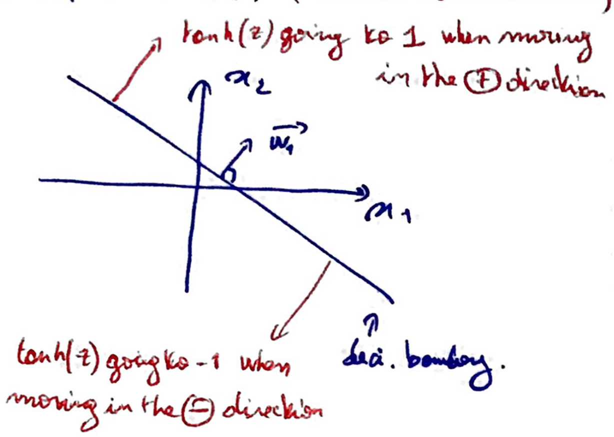




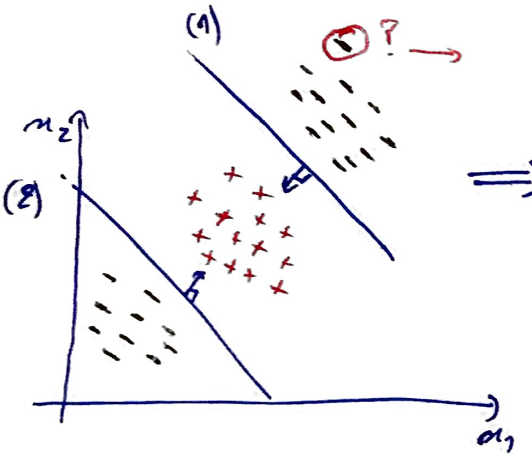
Neural signal transformation:



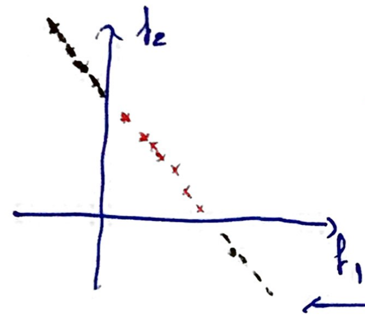
$z_1 = \vec{w_1} \cdot X + w_{01}$ (linear combination)



Hidden layer Unit:



Linear activation:



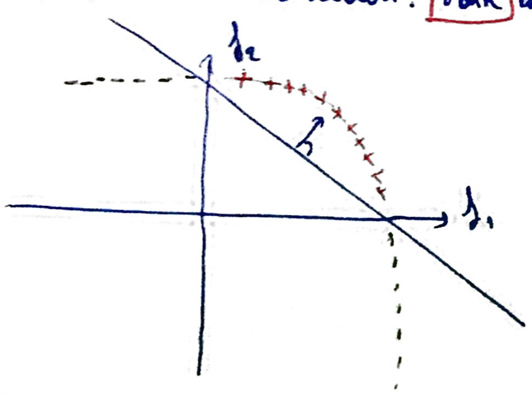
← It's not linearly separable.

For this point: $f_1 < 0$

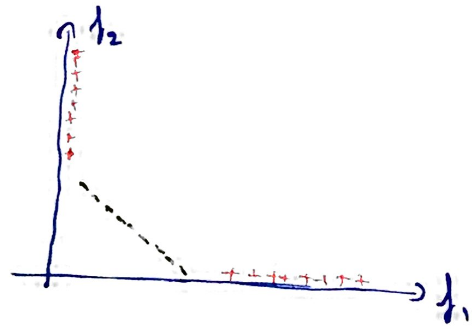
$f_2 \gg 0$

So we turn that linear activation unit into a non linear activation: $\boxed{\text{tanh}}$ activation

$\boxed{\text{Relu}}$ activation

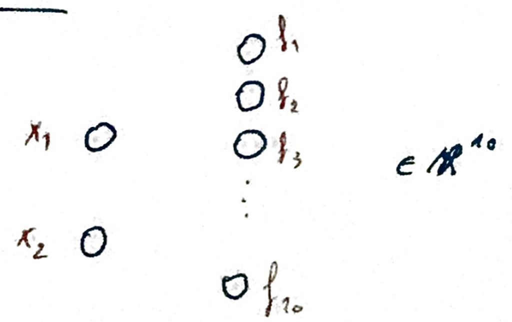


It's linearly separable.



It's not linearly separable.

Random Hidden unit:



Hidden units activation chosen at random

we would expect in that case that the problem might actually become separable in that new 10-dim space.

We find a 2-dim version of that, that highlights the points. The points are really separable now in that 10-dim space.

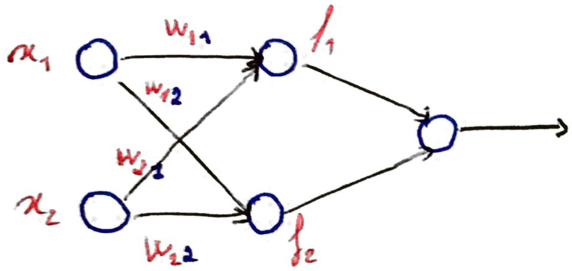
Introducing Redundancy will make the optimization problem that we have to solve easier.

let's consider a simple 2-dim classification task:

$$x^{(1)} = (-1,-1) \longrightarrow y^{(1)} = 1$$
$$x^{(2)} = (1,-1) \longrightarrow y^{(2)} = -1$$
$$x^{(3)} = (-1,1) \longrightarrow y^{(3)} = -1$$
$$x^{(4)} = (1,1) \longrightarrow y^{(4)} = 1$$

For simplicity, we are only interested in binary classification. (y can be 1 or -1)

linear Separability After first layer:



$$f_1^{(i)} = f\left(w_{01} + \left(w_{11} x_1^{(i)} + w_{21} x_2^{(i)}\right)\right)$$

$$f_2^{(i)} = f\left(w_{02} + \left(w_{12} x_1^{(i)} + w_{22} x_2^{(i)}\right)\right)$$

Consider $D' = \{([f_1^{(i)}, f_2^{(i)}], y^{(i)}),\ i = 1,2,3,4\}$ and $f(z) = 2z - 3$

Which of weights would the set D' be linearly separable?

ex: $w_{11} = w_{21} = 2$ | $w_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ $w_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$ $w_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

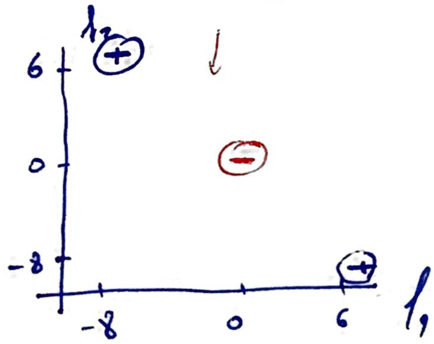
$w_{12} = w_{22} = -2$

$w_{01} = w_{02} = 1$

$$f_1^{(1)} = f\left(1 + \begin{bmatrix} 2 \\ 2 \end{bmatrix} [-1 \ -1]\right) = f(-3) = (2)(-3) - 3 = -9$$

$$f_2^{(1)} = f\left(1 + \begin{bmatrix} -2 \\ -2 \end{bmatrix} [-1 \ -1]\right) = f(5) = (2)(5) - 3 = 7$$

Not linearly separable.



Non-linear Activation Functions:

$$w_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad w_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad w_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

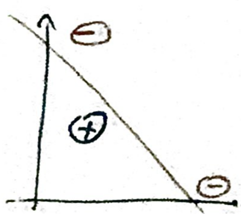
$f(z) = \mathrm{Relu}(z) = \max\{0, z\}$

$$f_1^{(1)} = f\left(1 + \begin{bmatrix} 1 \\ -1 \end{bmatrix} [-1 \ -1]\right) = f(+1) = \mathrm{Relu}(+1) =$$

$$f_2^{1} = f\left(1 + \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-1 \ -1]\right) = f(1) = \mathrm{Relu}(1) = 1$$

$\Rightarrow (1,1)\ (3,0)\ (0,3)\ (1,1)$

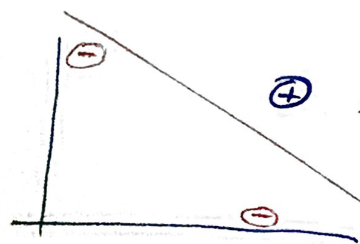
It's linearly separable



$$f(z) = \tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

$f_1^{1}(1) = \tanh(1) = 0.76$ ...

$f_2^{1}(1) = \tanh(1) = 0.76$



It's linearly separable