Mathematics in Deep Learning

Lachlan Jones

MATH199

2025

Mathematics in Deep Learning

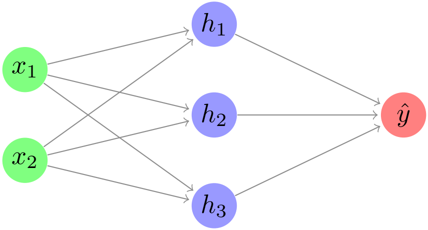


Figure: MLP network with two inputs and one hidden layer

Remark

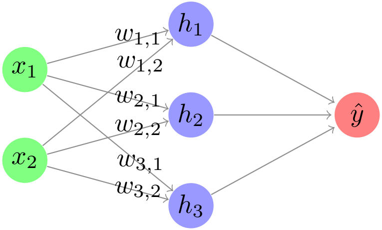
We've tidied up our diagram to not show biases and activation functions, but they're still there and being used in calculations.

Calculations:

$$h_1 = w_{1,1} \cdot x_1 + w_{1,2} \cdot x_2 + b_1$$

$$h_2 = w_{2,1} \cdot x_1 + w_{2,2} \cdot x_2 + b_2$$

$$h_3 = w_{3,1} \cdot x_1 + w_{3,2} \cdot x_2 + b_3$$

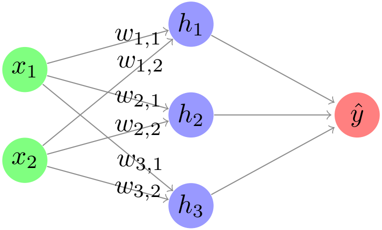


Calculations:

$$h_1 = w_{1,1} \cdot x_1 + w_{1,2} \cdot x_2 + b_1$$

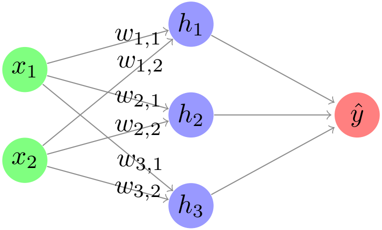
$$h_2 = w_{2,1} \cdot x_1 + w_{2,2} \cdot x_2 + b_2$$

$$h_3 = w_{3,1} \cdot x_1 + w_{3,2} \cdot x_2 + b_3$$



This is a typical system of equations, so we can tidy things up with matrix algebra:

Calculations:



$$h_1 = w_{1,1} \cdot x_1 + w_{1,2} \cdot x_2 + b_1$$

$$h_2 = w_{2,1} \cdot x_1 + w_{2,2} \cdot x_2 + b_2$$

$$h_3 = w_{3,1} \cdot x_1 + w_{3,2} \cdot x_2 + b_3$$

This is a typical system of equations, so we can tidy things up with matrix algebra:

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

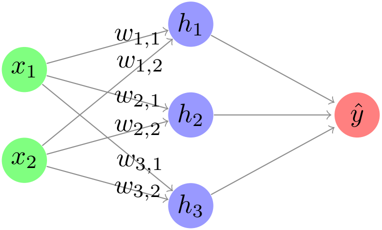
Mathematics in Deep Learning

Calculations:

$$h_1 = w_{1,1} \cdot x_1 + w_{1,2} \cdot x_2 + b_1$$

$$h_2 = w_{2,1} \cdot x_1 + w_{2,2} \cdot x_2 + b_2$$

$$h_3 = w_{3,1} \cdot x_1 + w_{3,2} \cdot x_2 + b_3$$



This is a typical system of equations, so we can tidy things up with matrix algebra:

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

With an activation function, the output of the hidden layer is:

$$\underline{h} = f(W \cdot \underline{x} + \underline{b})$$

Typically, architecture is much more complex, with many hidden layers. This is what deep learning is referring to.

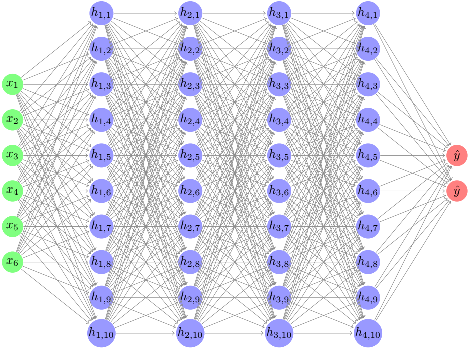


Figure: Example of what a deep NN could look like

Mathematics in Deep Learning

With all these different edges, deep neural networks have the capacity to learn complex, non-linear relationships between input data ($\underline{x}$) and output label ($\hat{y}$).

With all these different edges, deep neural networks have the capacity to learn complex, non-linear relationships between input data ($\underline{x}$) and output label ($\hat{y}$).

Suitable values for these edges need to be determined in a process called training, which uses an algorithm called gradient descent (very similar to Euler's Method, i.e. using calculus).

Some Examples:

- Using measurements of chemical modifications to human DNA to predict lifetime smoking exposure - this is what my Master's thesis is on

Mathematics in Deep Learning

Some Examples:

- Using measurements of chemical modifications to human DNA to predict lifetime smoking exposure - this is what my Master's thesis is on
- Teaching computers to recognise handwritten digits - `https://okdalto.github.io/VisualizeMNIST_web/`

Some Examples:

- Using measurements of chemical modifications to human DNA to predict lifetime smoking exposure - this is what my Master's thesis is on
- Teaching computers to recognise handwritten digits - `https://okdalto.github.io/VisualizeMNIST_web/`
- Predicting the most likely word to next appear in a sequence, aka large language models (ChatGPT)

Some Examples:

- Using measurements of chemical modifications to human DNA to predict lifetime smoking exposure - this is what my Master's thesis is on
- Teaching computers to recognise handwritten digits - `https://okdalto.github.io/VisualizeMNIST_web/`
- Predicting the most likely word to next appear in a sequence, aka large language models (ChatGPT)

Each of these use the same connections we saw before, just arranged in different ways to form different architectures.