2025-02-25T02-57-44
Tags: DLS

DLS C1 week 1

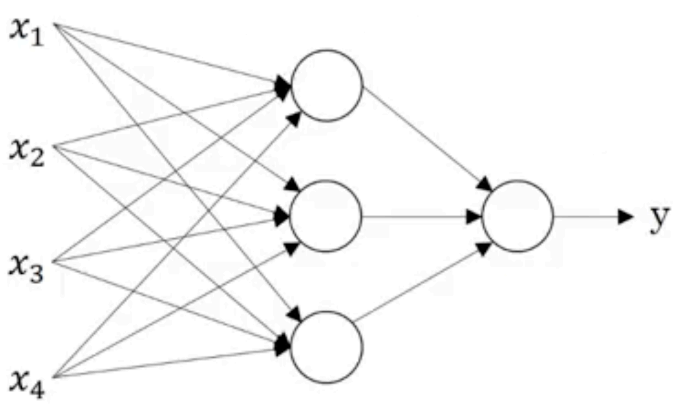
What is a Neural Network?

A function which fits the price of a house based on its size is a simple example of a neural network.

Here is a simple implementation of a neuron.

1. You input a size.
2. The model computes a linear function $y = w \cdot x + b$ where w is weight, x is input, and b is bias.
3. The model uses an activation function by taking a max of zero and the result—known as ReLU.
4. Obtain the output $\hat{y}$

Given enough training examples of (x, y), neural networks are great at figuring out accurate mapping functions from x to y.



The image above shows a neural network with an input layer with four input features i.e., $x = [x_1, x_2, x_3, x_4]$.

The neurons at the center of the figure are called *hidden units*—neurons that are not part of the input and output layer.
Explictly adding hidden units is not a necessity. But, adding them may help.

A *fully connected (hidden) layer* contains hidden units which receive inputs from all features.
While common, it is not a necessity for a hidden unit to take every feature.

Supervised learning

Supervised learning is where you have an input x and you want to learn a function that maps to some output y.

Types of neural networks:

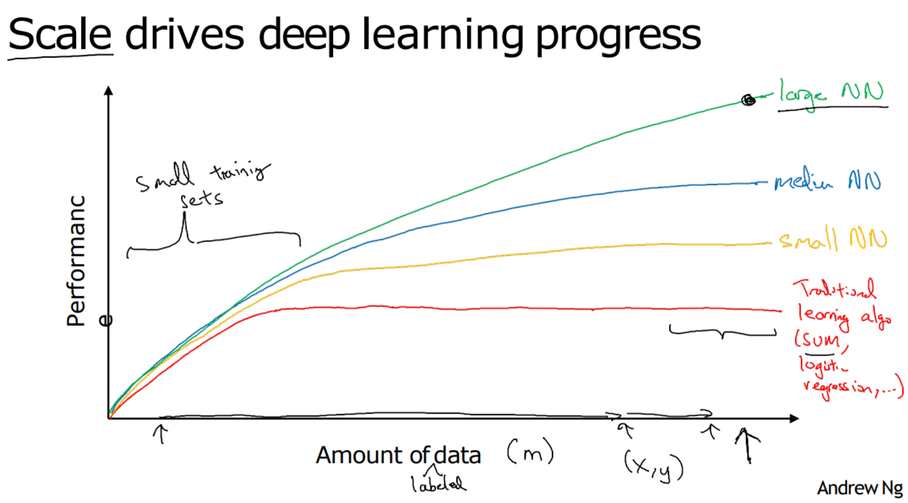
- Standard neural networks are for classical regression and classification.
- Convolutional neural networks are for spatial data like images.
- Recurrent neural networks are for sequential and temporal data like text and speech.

There's hybrid neural networks which are used for custom and complex prediction problems.

There exists two types of data.

- Structured data: each features are well-defined e.g., tabular data.
- Unstructured data: raw data e.g., audio, image, and text.

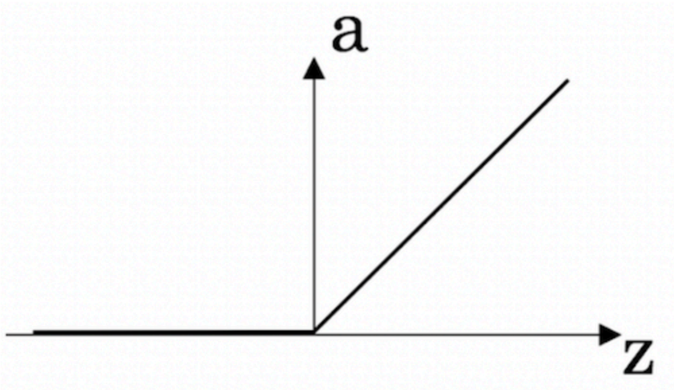
Scale



The performance of a neural network depend on:

- Amount of data.
- Size of the neural network.

Neural networks are better than traditional learning algorithms on large datasets—however, not on small datasets.
On small datasets, performance depends on hand-engineering features.



Algorithms are crafted to make neural networks run faster. Switching to ReLU instead of a sigmoid as the activation function is a prime example.

- A sigmoid function contains regions where the gradient is zero—slows down learning.
- ReLU makes the gradient less likely to shrink to zero by making the gradient equal to one for all positive values.