Deep Learning - Introduction

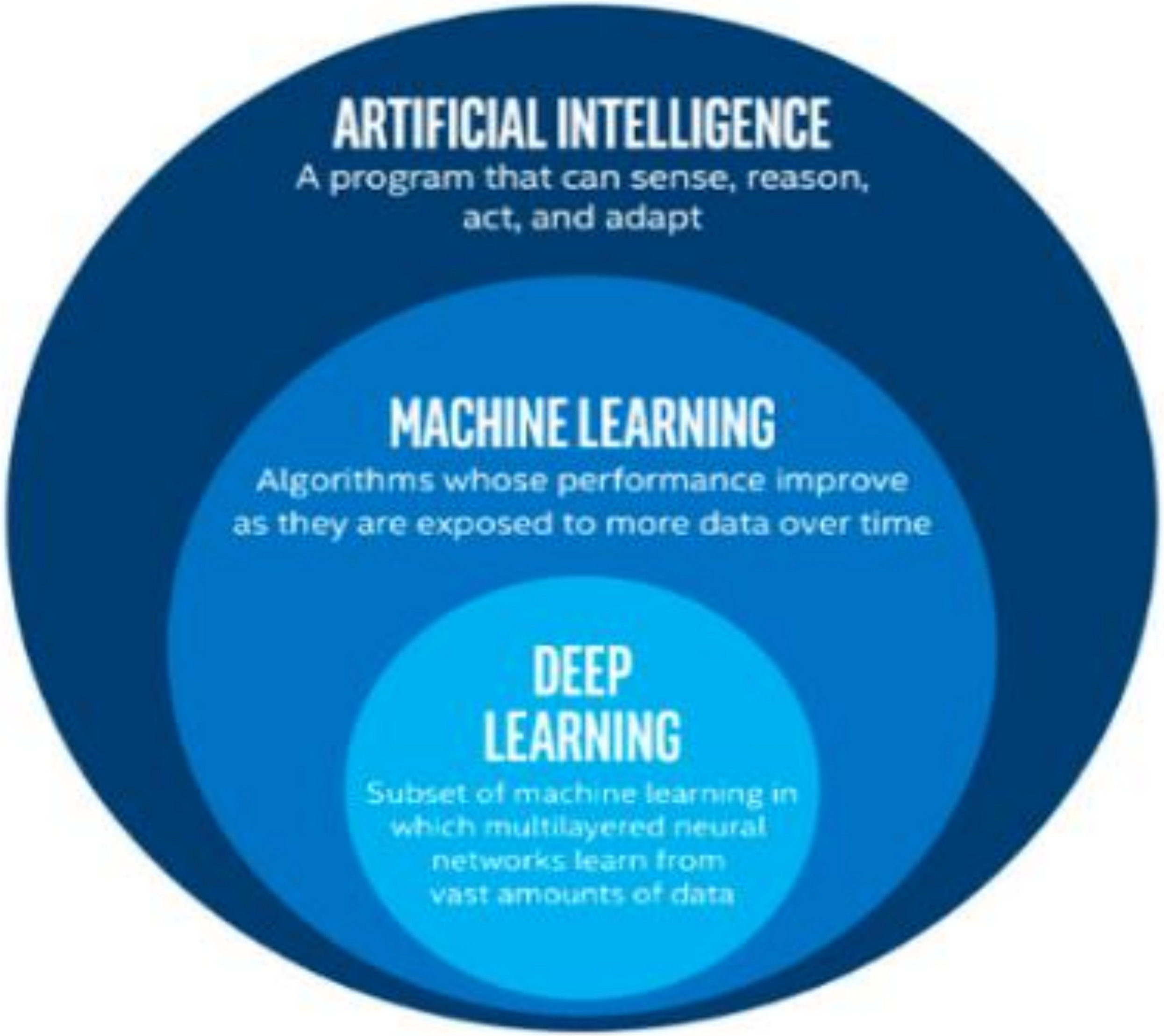
Deep Learning

- How Google can translate entire paragraphs from one language into another in a matter of milliseconds;
- How Netflix and YouTube can provide good recommendations;
- How self-driving cars are possible?
- All of these innovations are the product of deep learning and artificial neural networks.
- All recent advances in artificial intelligence in recent years are due to deep learning.
- Without deep learning, we would not have self-driving cars, chatbots or personal assistants like Alexa and Siri.
- Google Translate would continue to be as primitive as it was 10 years ago before Google switched to neural networks and Netflix would have no idea which movies to suggest.
- Neural networks are behind all these technologies

WHAT IS DEEP LEARNING AND HOW DOES IT WORK?

- Deep learning is a type of machine learning, inspired by the structure of the human brain.
- Deep learning algorithms attempt to draw similar conclusions as humans would by continually analyzing data with a given logical structure.
- To achieve this, deep learning uses multi-layered structures of algorithms called neural networks

AI vs. Machine Learning vs. Deep Learning



AI vs. Machine Learning vs. Deep Learning

- Artificial intelligence is a general term that refers to techniques that enable computers to mimic human behavior
- Machine learning represents a set of algorithms trained on data that make all of this possible
- Deep learning is a subset of machine learning, which is a subset of artificial intelligence
- Deep learning is a type of machine learning, inspired by the structure of the human brain

AI vs. Machine Learning vs. Deep Learning

- The design of the neural network is based on the structure of the human brain
- Just as brains identify patterns and classify different types of information, we can teach neural networks to perform the same tasks on data
- The individual layers of neural networks can also be thought of as a sort of filter that works from gross to subtle, which increases the likelihood of detecting and outputting a correct result
- The human brain works similarly
- Whenever we receive new information, the brain tries to compare it with known objects
- The same concept is also used by deep neural networks
- Neural networks enable us to perform many tasks, such as clustering, classification or regression

Advantage of deep learning over machine learning

Advantage of deep learning over machine learning

1. No feature extraction
2. Powered by massive amounts of data - Era of Big data

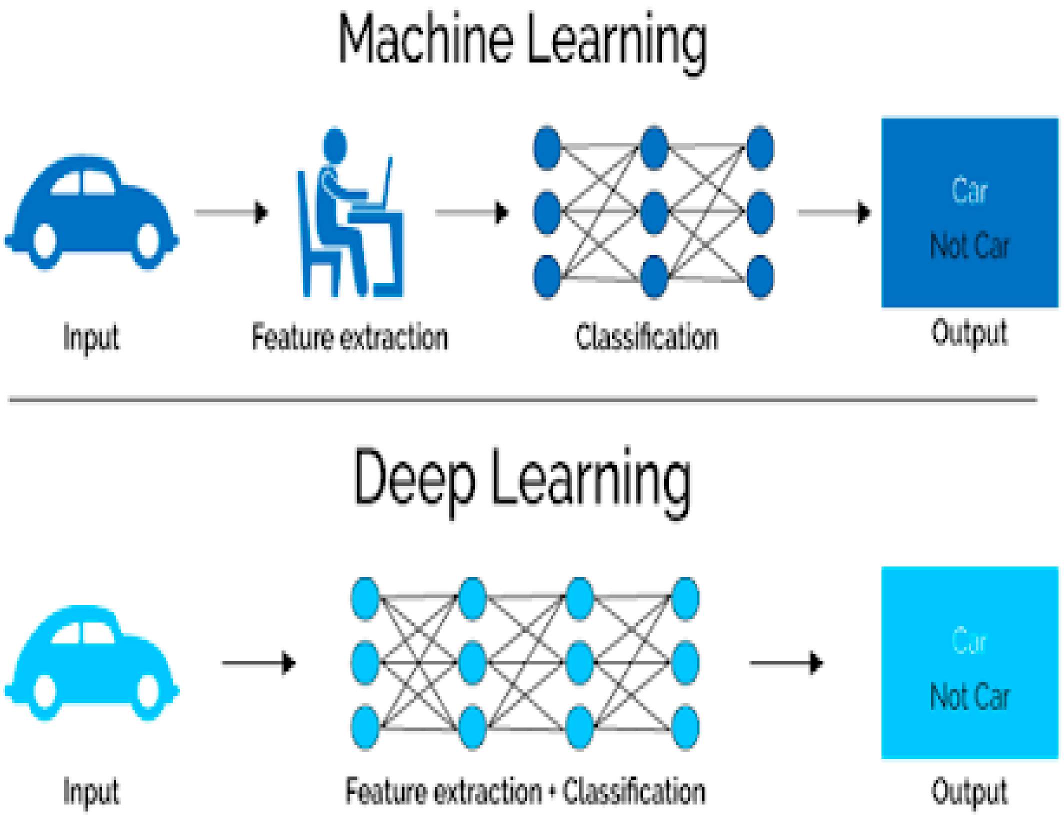
Advantage of deep learning over machine learning - NO FEATURE EXTRACTION

- The first advantage of deep learning over machine learning is the redundancy of the feature extraction.
- Traditional machine learning methods including decision trees, SVM, naïve Bayes classifier and logistic regression algorithms are called flat algorithms.
- “Flat” refers to the fact these algorithms cannot normally be applied directly to the raw data (such as .csv, images, text, etc.). We need a preprocessing step called feature extraction.
- Deep learning’s artificial neural networks do not need the feature extraction step.
- A more and more abstract and compressed representation of the raw data is produced over several layers of an artificial neural net.
- We then use this compressed representation of the input data to produce the result.
- we can say that the feature extraction step is already part of the process that takes place in an artificial neural network.
- During the training process, this neural network optimizes this step to obtain the best possible abstract representation of the input data.
- This means that deep learning models require little to no manual effort to perform and optimize the feature extraction process.

Feature Engineering

- Features are transformations on input data that facilitate a downstream algorithm, like a classifier, to produce correct outcomes on new data.
- Feature engineering consists of coming up with the right transformations so that the downstream algorithm can solve a task.
- For instance, in order to tell ones from zeros in images of handwritten digits, we would come up with a set of filters to estimate the direction of edges over the image, and then train a classifier to predict the correct digit given a distribution of edge directions.
- Another useful feature could be the number of enclosed holes, as seen in a zero, an eight, and, particularly, loopy twos

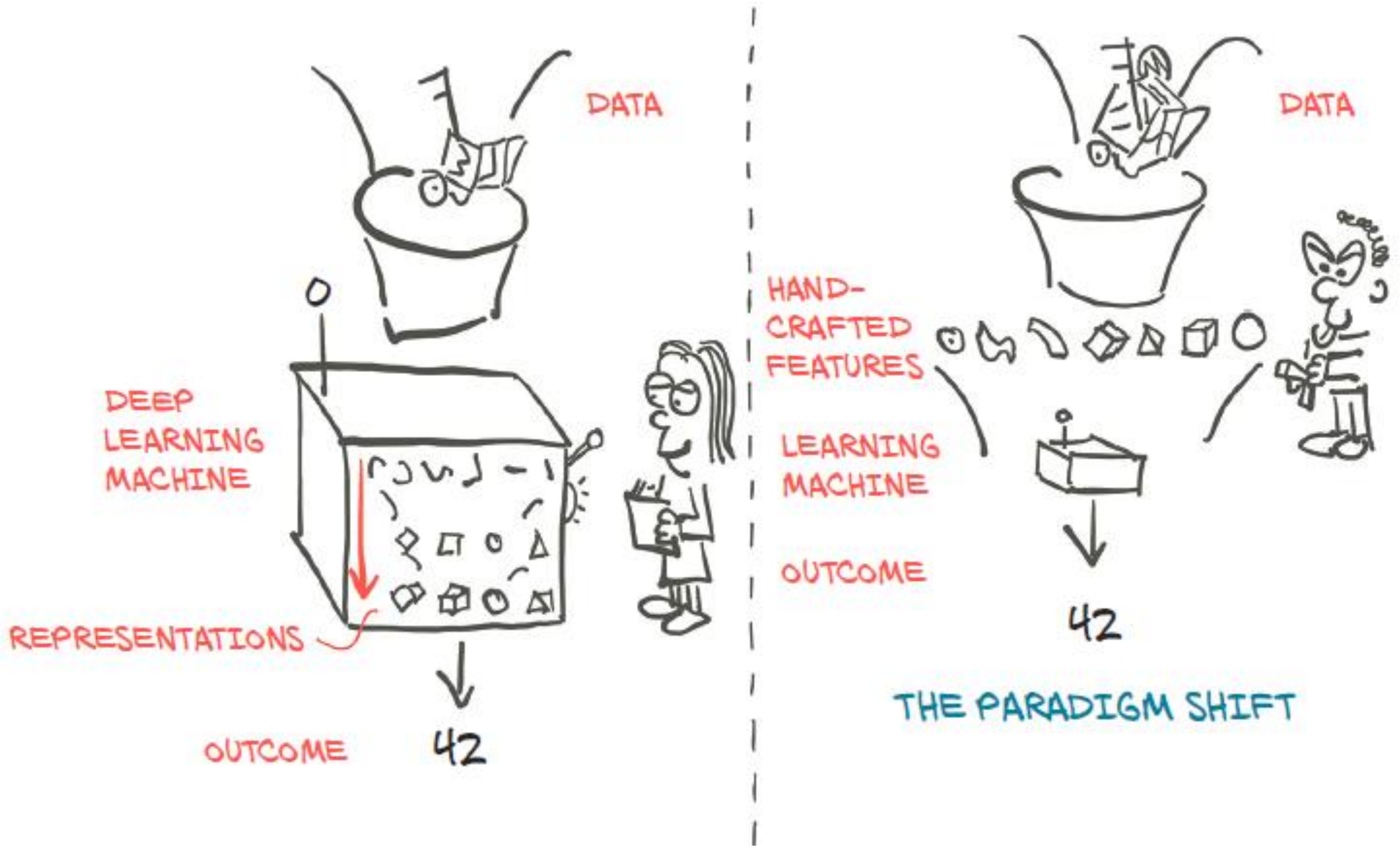
NO FEATURE EXTRACTION



How to use a machine learning model

- To determine if a particular image is showing a car or not
  - we humans first need to identify the unique features of a car (shape, size, windows, wheels, etc.)
  - then extract the feature and give it to the algorithm as input data.
  - In this way, the algorithm would perform a classification of the images.
- That is, in machine learning, a programmer must intervene directly in the action for the model to come to a conclusion.
- In the case of a deep learning model, the feature extraction step is completely unnecessary.
  - The model would recognize these unique characteristics of a car and make correct predictions without human intervention.

Deep Learning Machine



Deep learning exchanges the need to handcraft features for an increase in data and computational requirements

Deep Learning Machine

What we need to execute successful deep learning shown in the previous diagram:

- We need a way to ingest whatever data we have at hand.
- We somehow need to define the deep learning machine.
- We must have an automated way, training, to obtain useful representations and make the machine produce desired outputs.

Deep Learning Machine

- At its core, the deep learning machine shown in figure is a complex mathematical function mapping inputs to an output.
- To facilitate expressing this function, PyTorch provides a core data structure, the tensor
- Tensor is a multidimensional array that shares many similarities with NumPy arrays.
- PyTorch comes with features to perform accelerated mathematical operations on dedicated hardware
- This makes it convenient to design neural network architectures and train them on individual machines or parallel computing resources

PyTorch

- PyTorch is a high-performance library with optimization support for scientific computing in Python:

1. provides accelerated computation using Graphical Processing Units (GPUs), yielding speedups in the range of 50x over doing the same calculation on a CPU.

2. provides facilities that support numerical optimization on generic mathematical expressions, which deep learning uses for training

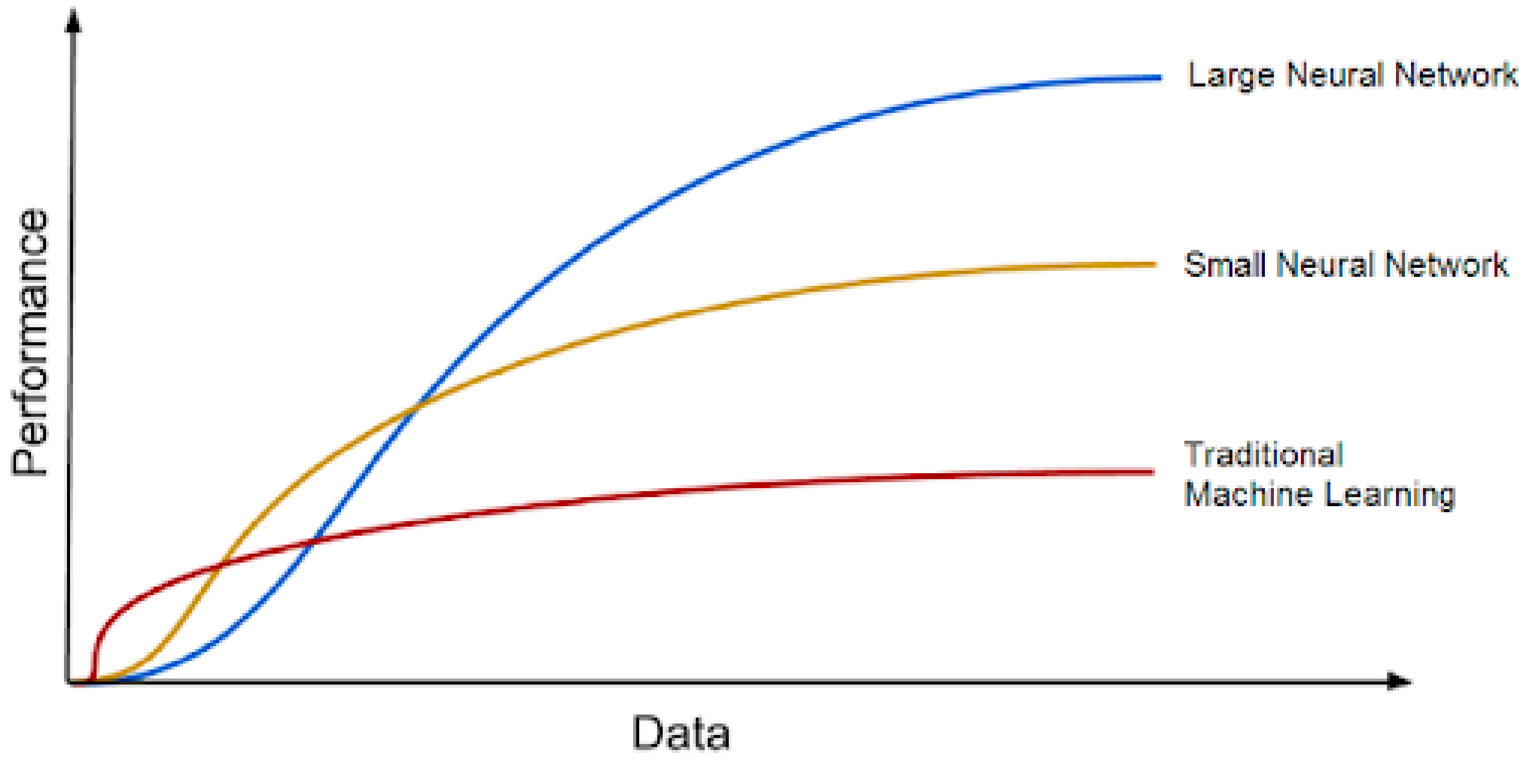
Advantage of deep learning over machine learning - THE ERA OF BIG DATA

- The second huge advantage of deep learning is that it is powered by massive amounts of data.
- The era of big data will provide huge opportunities for new innovations in deep learning
- Andrew Ng, the chief scientist of China's major search engine Baidu, co-founder of Coursera and one of the leaders of the Google Brain Project, puts it this way:

- *I think AI is akin to building a rocket ship. You need a huge engine and a lot of fuel. If you have a large engine and a tiny amount of fuel, you won't make it to orbit. If you have a tiny engine and a ton of fuel, you can't even lift off. To build a rocket you need a huge engine and a lot of fuel.*

- The analogy to deep learning is that the rocket engine is the deep learning models and the fuel is the huge amounts of data we can feed to these algorithms.

Deep learning algorithms improve with increasing amounts of data



Deep learning models tend to increase their accuracy with the increasing amount of training data, whereas traditional machine learning models such as SVM and naive Bayes classifier stop improving after a saturation point

Deep Learning

- Deep learning uses large amounts of data to approximate complex functions whose inputs and outputs are far apart
  - like an input image and, as output, a line of text describing the input;
  - or a written script as input and a natural-sounding voice reciting the script as output;
  - or, even more simply, associating an image of a golden retriever with a flag that tells us “Yes, a golden retriever is present.”

DL History

- Concepts of neural networks and deep learning go back a long way.
- The proof that such a network could function as a way of replacing any mathematical function in an approximate way,
- which underpins the idea that neural networks can be trained for many different tasks, dates back to 1989
- Convolutional Neural Networks (CNNs) were being used to recognize digits on check in the late ’90s.

Reasons for DL popularity

- Surge in Graphical Processing Units (GPUs) performance and their increasing affordability.
- Designed originally for gaming, GPUs need to perform countless millions of matrix operations per second
- graphics cards could be used to speed up training as well as make larger, deeper neural network architectures feasible for the first time.
- Techniques such as Dropout not just speed up training but make training more generalized (so that the network does not just learn to recognize the training data, a problem called overfitting)
- Companies have taken GPU-based approach to the next level, with Google creating Tensor Processing Units (TPUs), which are devices custom-built for performing deep learning as fast as possible, and are even available to the general public as part of their Google Cloud ecosystem.

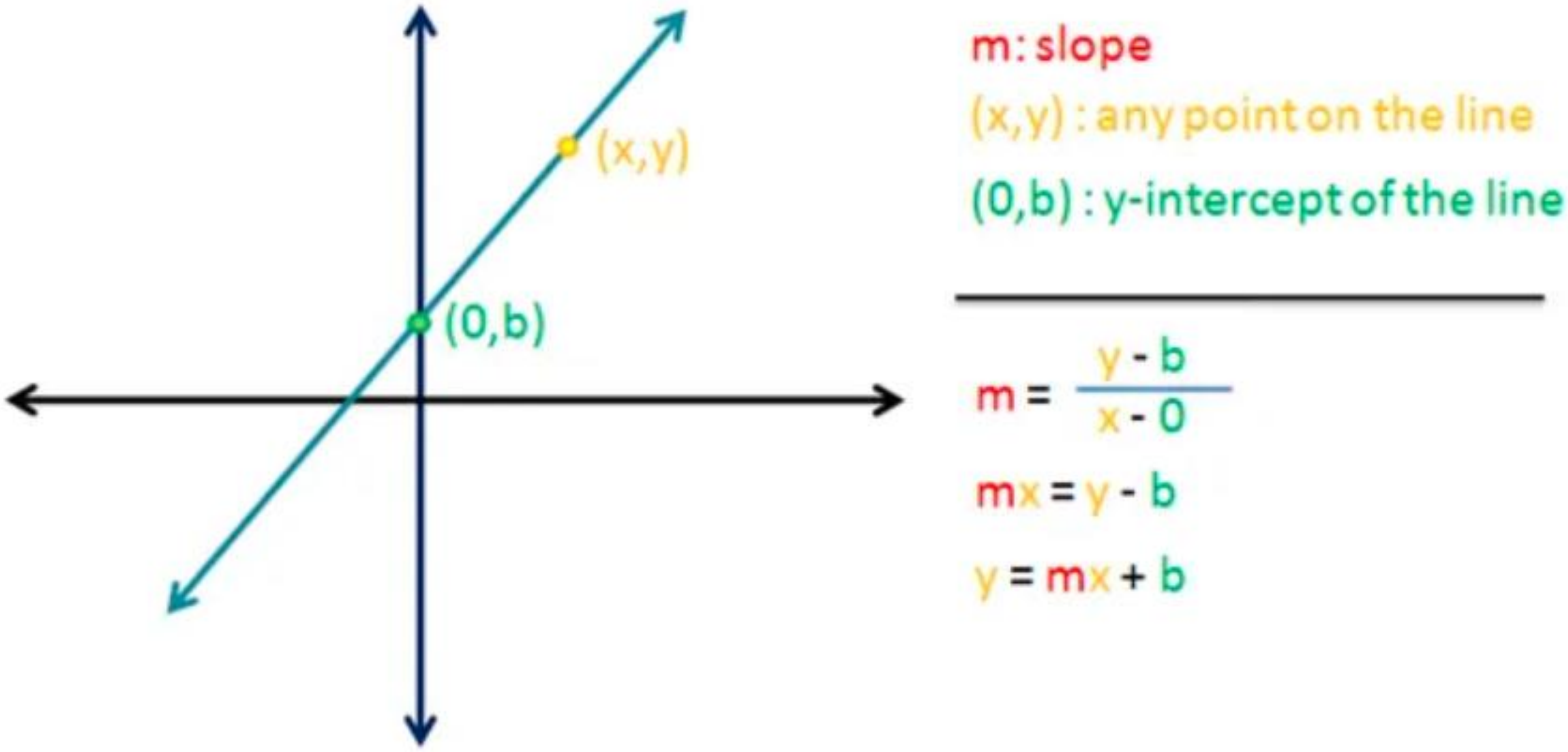
Deep Learning - Meaning

- “Deep learning” - Synonyms for neural networks
- Emphasizes that the algorithms often involve hierarchies with many stages of processing

Linear vs nonlinear

- We already have learning algorithms like linear regression, logistic regression, decision trees and random forests
- Linear regression assumes a linear relationship between the independent variables and the target variable.
- It seeks to fit a straight line or hyperplane that best represents the relationship between the variables.
- However, many real-world phenomena exhibit complex nonlinear patterns and interactions that cannot be accurately modeled by a simple linear relationship.
- This limitation arises from the underlying assumptions and structure of linear regression models
- Linear regression falls short in effectively capturing complex nonlinear relationships, but neural networks excel in this aspect

Linear Equation plotted on a graph



The Linear Regression Equation

A single neuron in the neural network works as a straight line which has the following equation:

This is the fundamental equation around which the whole concept of neural networks is based on

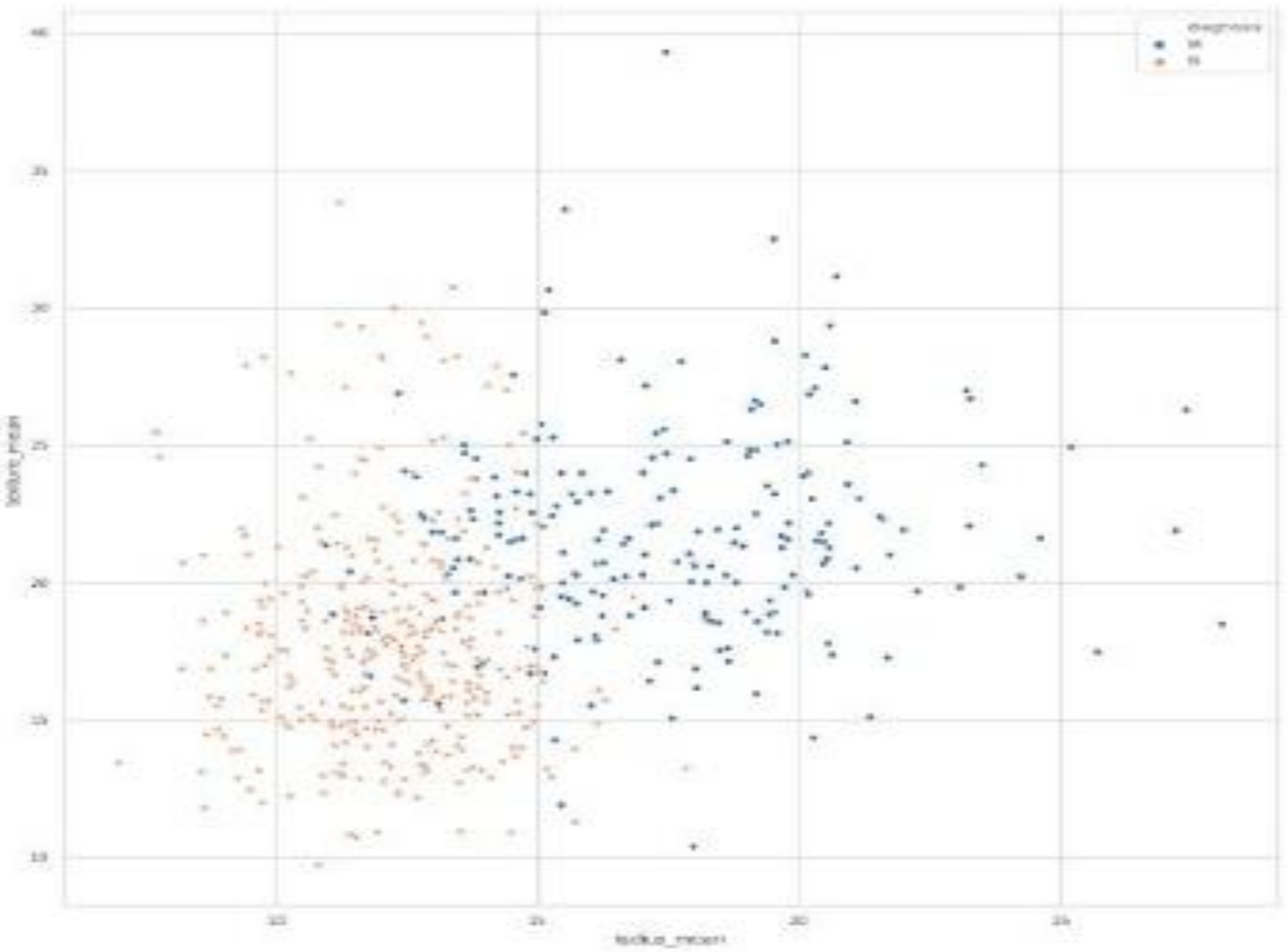
y: Dependent variable (Output of the neural network)

m: Slope of the line

x: Independent variable (Input features)

b: y-intercept

A complex, non-linear hypothesis – Plot of Cancer from Wisconsin data set



Only 2 features taken out of 30

- ❖ mean radius
- ❖ mean texture
- to gain information whether the tumor is Malignant (M) or Benign (B)
- M, represented by blue dots
- B, represented by an orange x

Non-linear hypothesis

- This is a supervised learning classification problem.
- If we are to apply logistic regression to this problem, the hypothesis would look

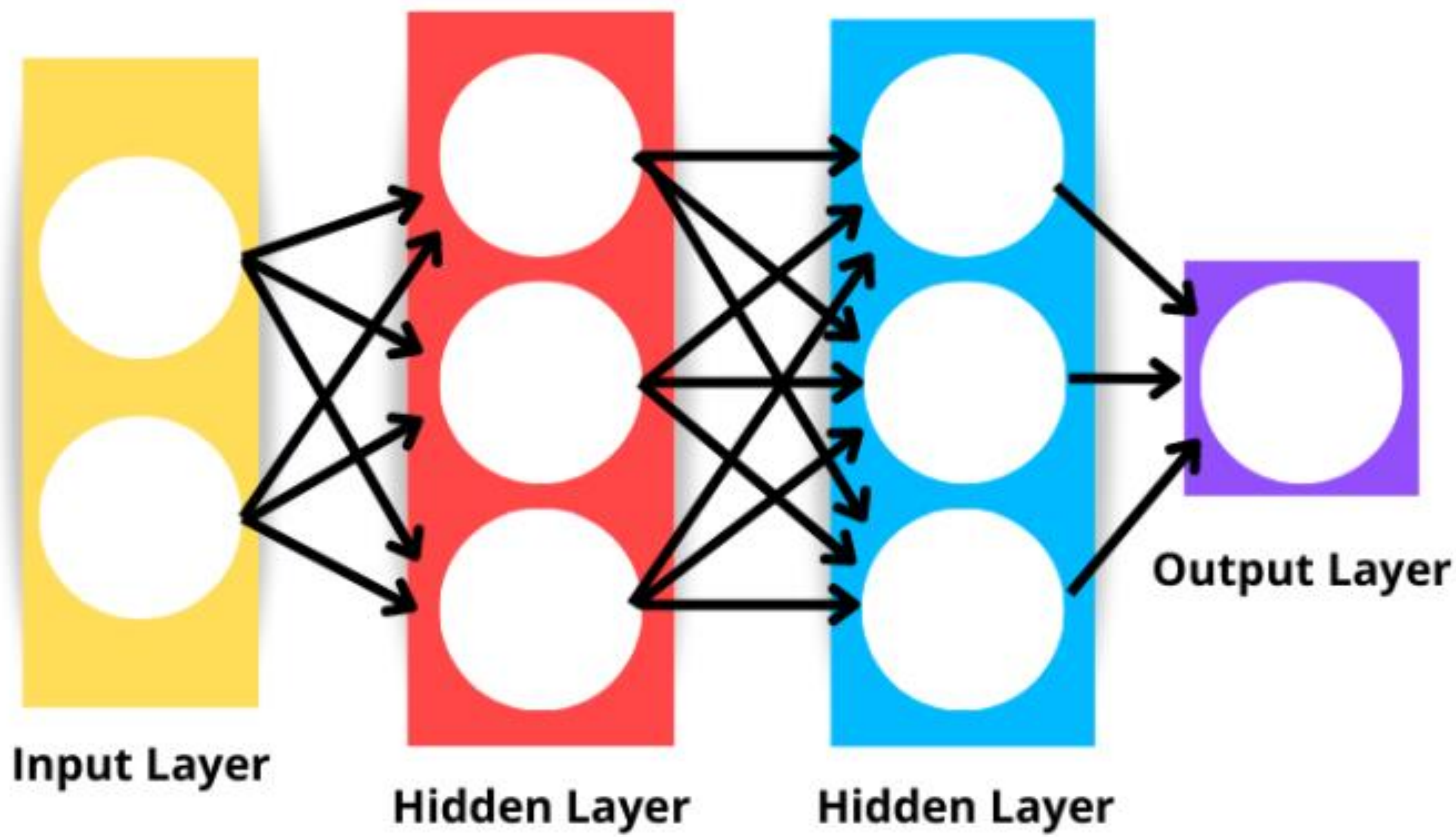
$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 {x_1}^2 x_2 + \theta_5 {x_1}^3 x_2 + \theta_6 x_1 {x_2}^2 + \cdots)$$

- g is the sigmoid function.
- If we perform logistic regression with such a hypothesis, then we might get a boundary — an extremely curvy one — separating the malignant and benign tumors
- But this is effective only when we consider two features
- But this data set contains 30 features

Non-linear hypothesis

- If we were to include only the quadratic terms in the hypothesis, we still have hundreds of non-linear features.
- The number of quadratic features generated will be to the order of $O(n^2)$, where n is the number of features (30).
- We end up overfitting the data set.
- It is also computationally expensive to work with that many features.
- This is if we are only using quadratic terms
- But there are also cubic terms generated in a similar manner

A Neural Network



A Neural Network

- Circles are referred to as neurons
- These neurons are mathematical functions which, when given some input, generate an output.
- The output of neurons depends on the input and the parameters of the neurons.
- We can update these parameters to get a desired value out of the network.
- Each of these neurons can be defined using activation function
- A sigmoid function gives an output between zero to one for every input it gets.
- These sigmoid units are connected to each other to form a neural network

Linear vs nonlinear

Neural networks enhance linear regression in three significant ways:

1. Nonlinear Transformation: Unlike linear regression, neural networks apply nonlinear transformations on top of the linear transformation. This enables them to model and capture intricate nonlinear patterns in the data.
2. Multiple Layers: Neural networks consist of multiple layers, allowing them to capture interactions and dependencies between features. Each layer contributes to extracting higher-level representations of the input data, enabling more sophisticated modeling.
3. Multiple Hidden Units: Within each layer of a neural network, there are multiple hidden units. Each hidden unit performs its unique combination of linear and nonlinear transformations, providing flexibility in capturing complex relationships and enhancing the model's ability to learn intricate patterns.