Deep Learning – A gentle introduction

Instructor, Nero Chan Zhen Yu

Forward School

When we think about AI...

Adapted from UC Berkeley CS188 Lecture 1 Slides

What experts say about AI?

- *"The science and engineering of making intelligent machines, especially intelligent computer programs"*, father of Artificial Intelligence, John McCarthy

- *"AI is all about figuring what to do when you don't know what to do."*, Peter Norvig.

- To generalize:
  - *Basically, artificial intelligence (AI) is the ability of a machine or a computer program to think and learn. The concept of AI is based on the idea of building machines capable of thinking, acting, and learning like humans.*
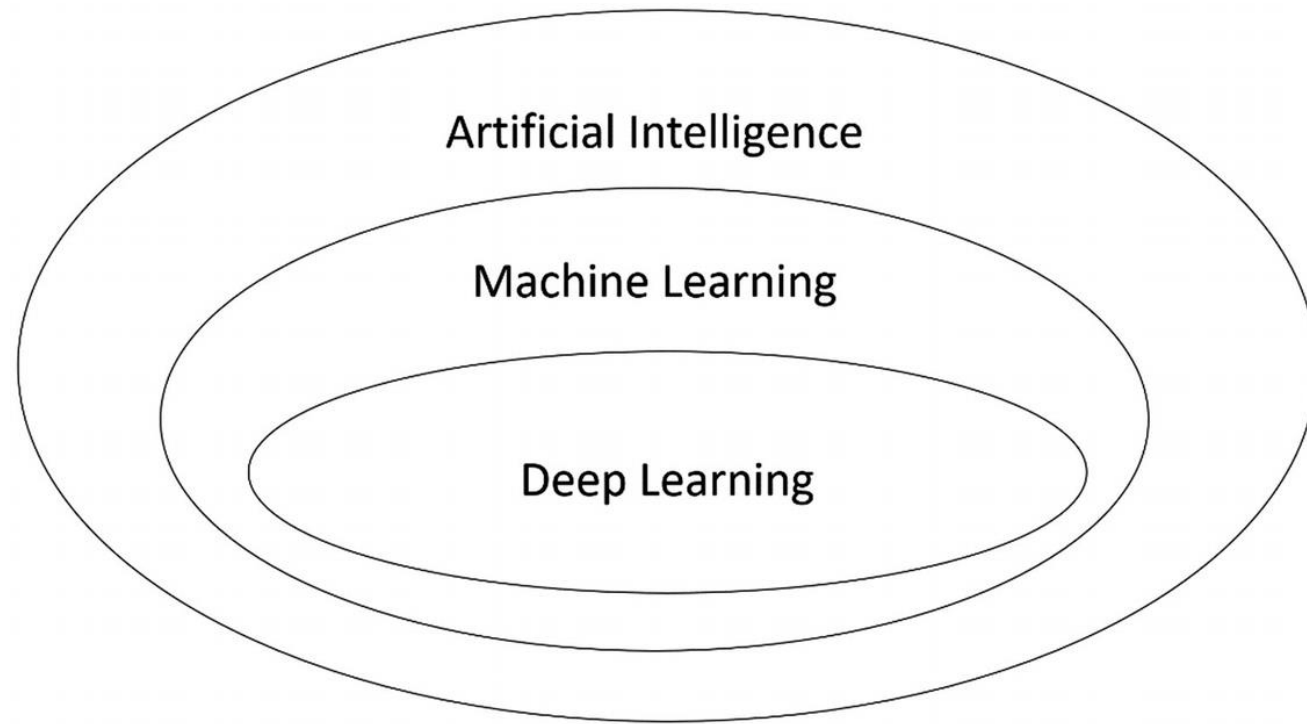
Structure of AI

Image from: Figure 1-3 Artificial Intelligence Basics: A Non-technical Introduction
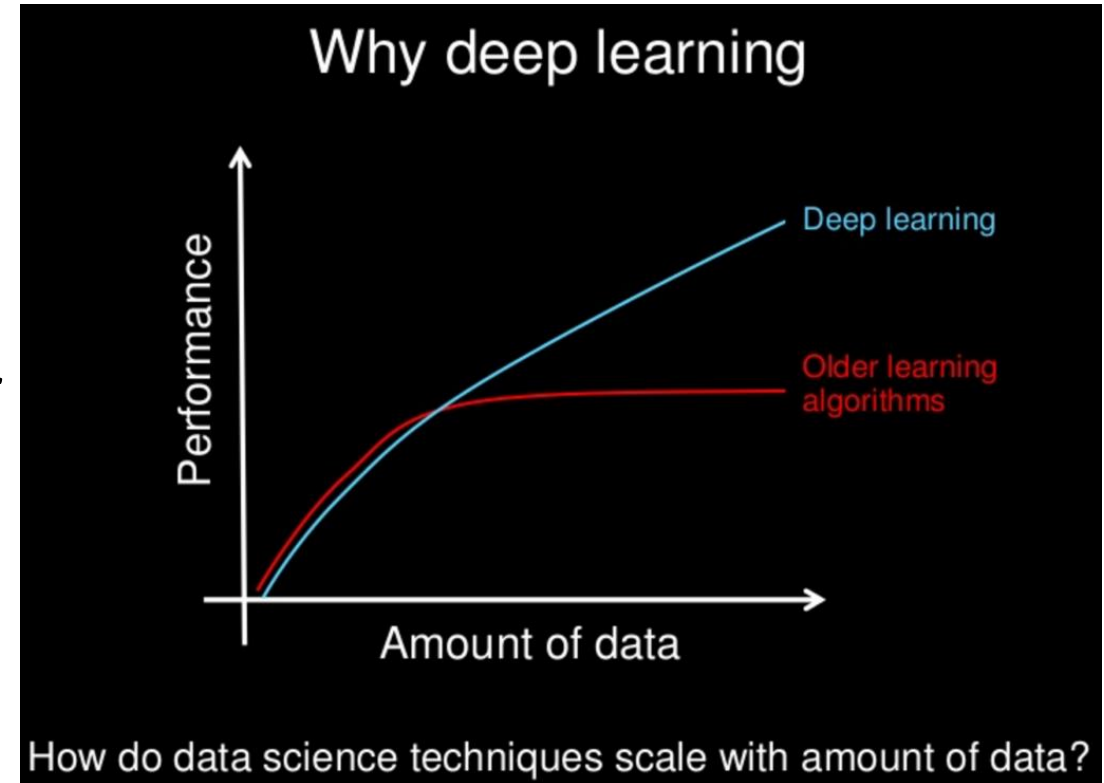
Deep Learning…

- A subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

- What's the differences between DL and (A)NN?

Forward School

DL (and/vs) NN…

- Andrew Ng:
  - *Using brain simulations, hope to:*
  - *– Make learning algorithms much better and easier to use.*
  - *– Make revolutionary advances in machine learning and AI.*
  - *I believe this is our best shot at progress towards real AI*
- Why now? – according to Andrew
  - We have more (labelled) data (in the scale of >1000 times)
  - We have fast enough computers to train the models
  - We can actually train large neural networks

Forward School

So what, then?

- *for most flavors of the old generations of learning algorithms … performance will plateau. … deep learning … is the first class of algorithms … that is scalable. … performance just keeps getting better as you feed them more data*

- *More importantly:*
  - *"one reason that deep learning has taken off like crazy is because it is fantastic at supervised learning"*

Why deep learning

Deep learning

Older learning algorithms

Performance

Amount of data

How do data science techniques scale with amount of data?

Forward School

Scalability of neural networks

Important Property of Neural Networks

Results get better with

more data +

bigger models +

more computation

(Better algorithms, new insights and improved techniques always help, too!)

Forward School

Deep Learning is

- Hierarchical Feature Learning
  - DL has the ability to perform automatic feature extraction from raw data, also called feature learning
- Yoshua Bengio describes deep learning in terms of the algorithms ability to discover and learn good representations using feature learning
  - Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features

Deep Learning is Neural Network ☺

Forward School

Why Deep Learning and not just call it Artificial Neural Networks?

- Because at the beginning, "Deep" was used to describe the development of LARGE artificial neural networks
  - Deep = many layered network

- Geoffrey Hinton
  - *Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory.*
  - *We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.*

Forward School

2016 talk by Geoffrey to the Royal Society

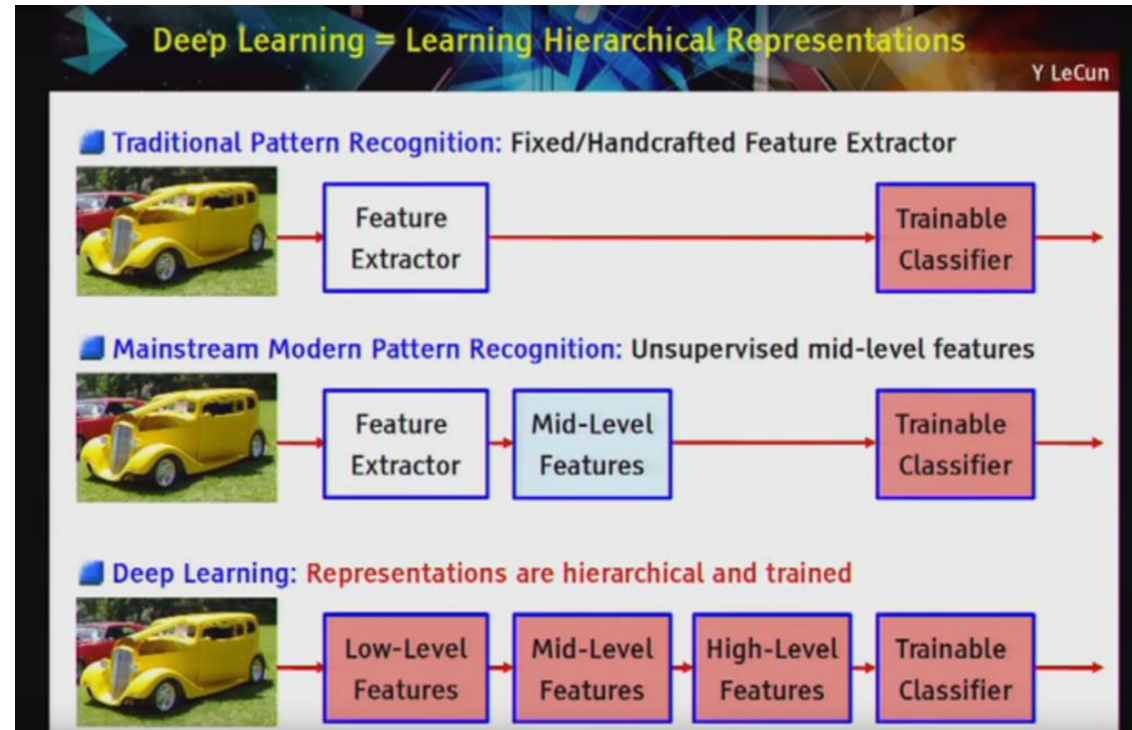What was actually wrong with backpropagation in 1986?

- We all drew the wrong conclusions about why it failed. The real reasons were:

1. Our labeled datasets were thousands of times too small.
2. Our computers were millions of times too slow.
3. We initialized the weights in a stupid way.
4. We used the wrong type of non-linearity.

Forward School

Then why Deep Learning is still getting the attention?

- Deep learning excels on problem domains where the inputs (and even output) are analog.
  - Meaning, they are not a few quantities in a tabular format but instead are images of pixel data, documents of text data or files of audio data.
- These are techniques used recently for surprising highly accurate object/face detection, autonomous driving and medical image recognition etc.

Forward School

Yann LeCun

- *deep learning [is] ... a pipeline of modules all of which are trainable. ... deep because [has] multiple stages in the process of recognizing an object and all of those stages are part of the training"*

Deep Learning = Learning Hierarchical Representations

Y LeCun

Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor

Feature Extractor → Trainable Classifier

Mainstream Modern Pattern Recognition: Unsupervised mid-level features

Feature Extractor → Mid-Level Features → Trainable Classifier

Deep Learning: Representations are hierarchical and trained

Low-Level Features → Mid-Level Features → High-Level Features → Trainable Classifier

Yann LeCun, Yoshua Bengio and Geoffrey Hinton in their Nature paper "Deep Learning"

- *Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.*

- *Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. [...] The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.*

- https://www.nature.com/articles/nature14539

Forward School

Summary

- For a layman explanation:
  - Deep learning is just very big neural networks on a lot more data, requiring bigger computers.

- Some popular techniques to begin with:
  - Multilayer Perceptron Networks.
  - Convolutional Neural Networks.
  - Long Short-Term Memory Recurrent Neural Networks.

Forward School