

Deep Learning

00 Introduction and Course logistics

Dr. Konda Reddy Mopuri
Dept. of AI, IIT Hyderabad
Jan-May 2025

Time slot

- B slot

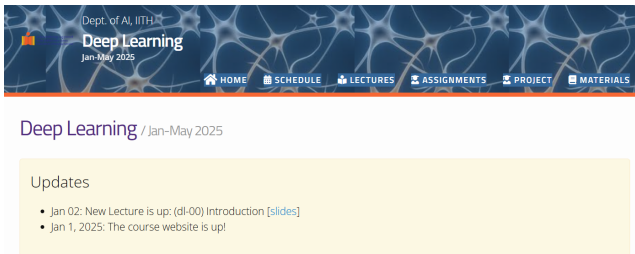
Time slot

- B slot
- Monday 10 - 10:55 AM
- Wednesday 9 - 9:55 AM
- Thursday 11 - 11:55 AM

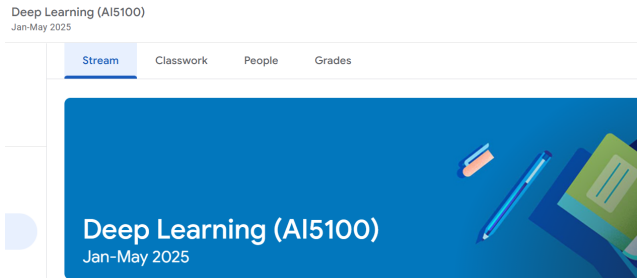
Time slot

- B slot
- Monday 10 - 10:55 AM
- Wednesday 9 - 9:55 AM
- Thursday 11 - 11:55 AM
- LHC-7

- Course website: <https://krmopuri.github.io/dl25/>



- Google Classroom (AI5100): ewt7tnh



- Joint the MS Teams (CS5480)

Welcome to Deep Learning (CS5480) - Jan 2025

Choose where you want to start



Upload Class Materials



Set up Class Notebook

- Programming Assignments - 30% (best 3 of 4)

- Programming Assignments - 30% (best 3 of 4)
- Written exams (best 3 of 4; pre-scheduled) - 30%

- Programming Assignments - 30% (best 3 of 4)
- Written exams (best 3 of 4; pre-scheduled) - 30%
- Kaggle Project (presentation) - 20%

- Programming Assignments - 30% (best 3 of 4)
- Written exams (best 3 of 4; pre-scheduled) - 30%
- Kaggle Project (presentation) - 20%
- Endsem (written) - 20%

- Schedule will be out on 6th Jan 2025

- Schedule will be out on 6th Jan 2025
- Mark your calendar - **No makeup tests! No partial penalties! No shifting the exam dates!**

- Sunayna Padhye (ai23mtech12002@iith.ac.in)
- Naveen George (ai23mtech12001@iith.ac.in)
- Madhumitha V (ai23resch11004@iith.ac.in)
- Harsh Udai (ai23resch11003@iith.ac.in)
- S S Madarkar (ai23mtech13003@iith.ac.in)
- Anish Pawar (ai23mtech14002)
- Puneet Rajan (ai23mtech11010)
- Riya Ann Easow (ee21btech11044@iith.ac.in)
- Some more will be added soon.

Deep Learning Contents

- Foundational ideas

Deep Learning Contents

- Foundational ideas
- Key concepts of modern architectures and techniques

Deep Learning (AI5100) Course Contents

Starting from an artificial neuron model, the aim of this course is to understand feed-forward, recurrent architectures of Artificial Neural Networks, all the way to the latest Generative AI models driven by Deep Neural Networks. Specifically, we will discuss the basic Neuron models (McCulloch Pitts, Perceptron), Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN, LSTM and GRU). We will understand these models' representational ability and how to train them using the Gradient Descent technique using the Backpropagation algorithm. We will then discuss the encoder-decoder architecture, attention mechanism and its variants. That will be followed by self-attention and Transformers. The next part of the course will be on Generative AI, wherein we will discuss Variational Autoencoders, GANs, Diffusion Models, GPT, BERT, etc. We will briefly discuss multi-modal representation learning (e.g., CLIP). Towards the end, students will be briefly exposed to some of the advanced topics and/or recent trends in deep learning.

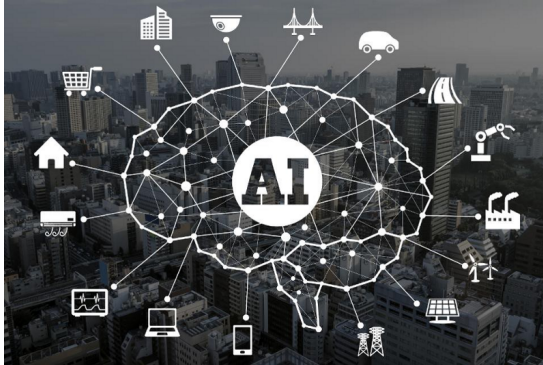
Prerequisites

- Programming in Python (**Primer on PyTorch on 18 January, online**)

Prerequisites

- Programming in Python (**Primer on PyTorch on 18 January, online**)
- A course on Machine Learning

Why Deep Learning?

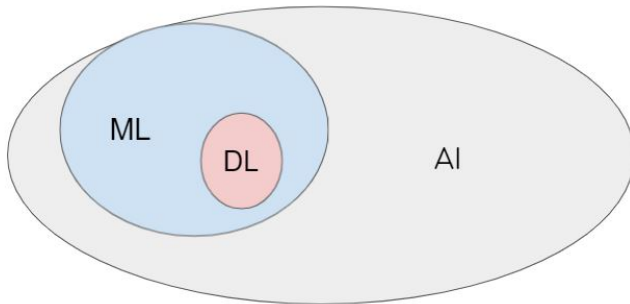


Deep Learning drives the recent AI boom. Image Source: Artificial Intelligence Magazine

- Lot of online resources
 - Deep Learning Foundations and Concepts by C Bishop & H Bishop
 - Michael Nielsen's text book on NN & DL
 - NPTEL course on Deep Learning by Prof. Mitesh Khapra, IITM
 - DL course by François Fleuret, Uni. of Geneva
 - Deep Learning textbook by Ian Goodfellow *et al.*
 - PyTorch - <https://pytorch.org/>
 - Many more that I could not list and am not aware of...

What is DL?

What is DL?



Subset of ML that is essentially Artificial Neural Networks with more layers

What is DL?

- Crude attempt to imitate the human brain in learning

Classical ML vs. DL

- Classical ML: Handcrafted features + learnable model

Classical ML vs. DL

- Classical ML: Handcrafted features + learnable model
- Need strong domain expertise

Classical ML vs. DL

- Classical ML: Handcrafted features + learnable model
- Need strong domain expertise

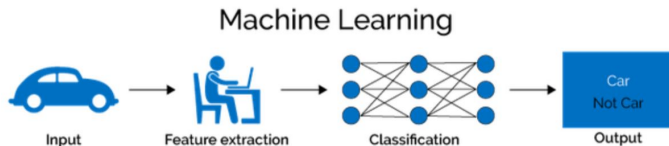


Figure credits: taken from [Jay Shaw in Quora](#) (not sure of authenticity)

Classical ML vs. DL

- Deep Learning: Deep stack of parameterized processing

Classical ML vs. DL

- Deep Learning: Deep stack of parameterized processing
- End-to-End learning

Classical ML vs. DL

- Deep Learning: Deep stack of parameterized processing
- End-to-End learning

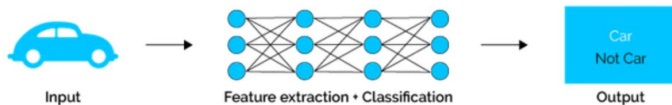


Figure credits: taken from [Jay Shaw in Quora](#) (not sure of authenticity)

Classical ML vs. DL

- ANNs predate some of the classical ML techniques

Classical ML vs. DL

- ANNs predate some of the classical ML techniques
- We are now dealing with a new generation ANNs

The Biological Neuron

- About 90 billion neurons in human brain

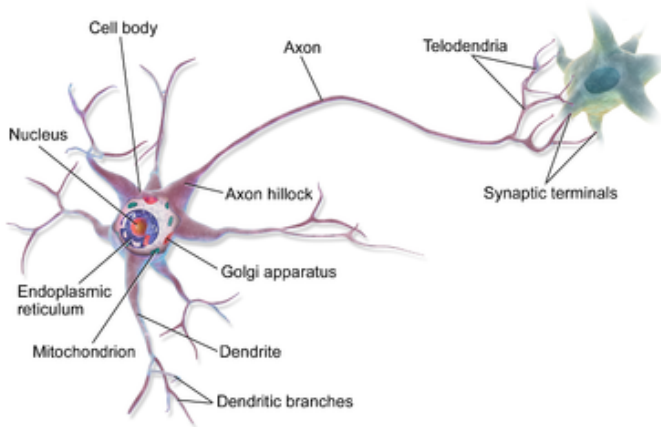


Figure credits: [Wikipedia](#)

① McCulloch Pitts neuron (1943) - Threshold Logic Unit

History of Neural Networks

- ① McCulloch Pitts neuron (1943) - Threshold Logic Unit
- ② Donald Hebb (1949) - Hebbian Learning Principle

History of Neural Networks

- ① McCulloch Pitts neuron (1943) - Threshold Logic Unit
- ② Donald Hebb (1949) - Hebbian Learning Principle
- ③ Marvin Minsky (1951) - created the first ANN (Hebbian Learning, 40 neurons)

History of Neural Networks

- ① McCulloch Pitts neuron (1943) - Threshold Logic Unit
- ② Donald Hebb (1949) - Hebbian Learning Principle
- ③ Marvin Minsky (1951) - created the first ANN (Hebbian Learning, 40 neurons)
- ④ Frank Rosenblatt (1958) - created perceptron to classify 20X20 images

History of Neural Networks

- ① McCulloch Pitts neuron (1943) - Threshold Logic Unit
- ② Donald Hebb (1949) - Hebbian Learning Principle
- ③ Marvin Minsky (1951) - created the first ANN (Hebbian Learning, 40 neurons)
- ④ Frank Rosenblatt (1958) - created perceptron to classify 20X20 images
- ⑤ David H Hubel and Torsten Wiesel (1959) demonstrated orientation selectivity and columnar organization in cats visual cortex

Backpropagation

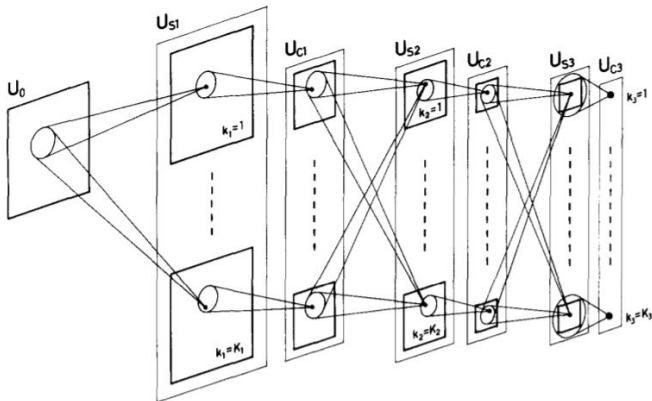
- Paul Werbos (1982) is credited for back-propagation in ANNs

Backpropagation

- Paul Werbos (1982) is credited for back-propagation in ANNs
- Precursors of BP were known as early as 1960s ([reference](#))

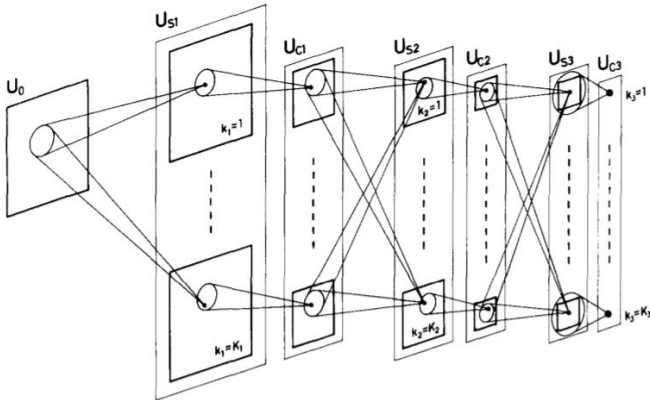
History (contd.)

① Neocognitron by Fukushima (1980)



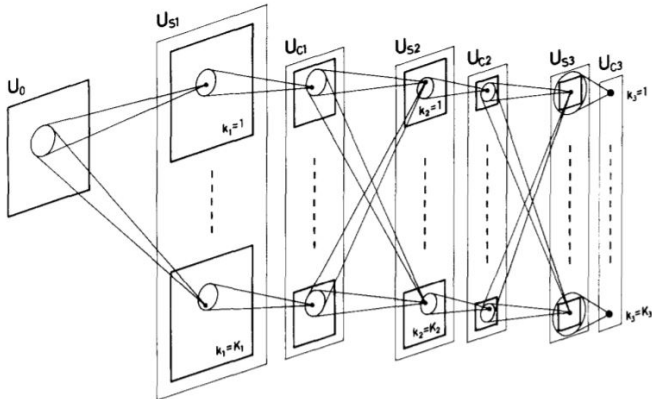
History (contd.)

- ① Neocognitron by Fukushima (1980)
- ② Implements the Hubel and Wiesel's principles



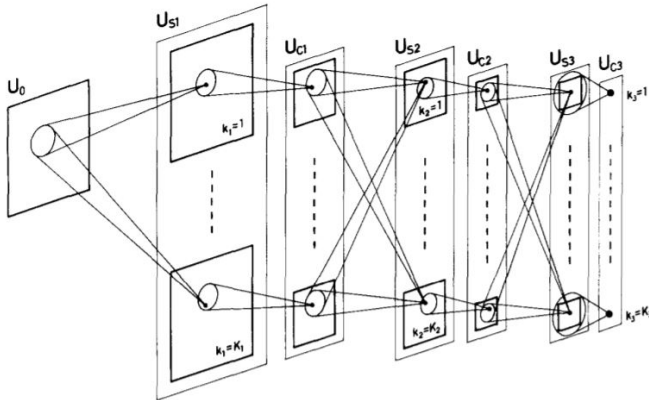
History (contd.)

- ① Neocognitron by Fukushima (1980)
- ② Implements the Hubel and Wiesel's principles
- ③ Used for hand-written digit recognition



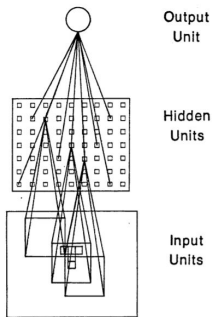
History (contd.)

- ① Neocognitron by Fukushima (1980)
- ② Implements the Hubel and Wiesel's principles
- ③ Used for hand-written digit recognition
- ④ Viewed as precursor for the modern CNNs



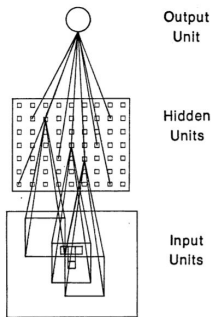
History (contd.)

① Rumelhart (1986) trained with backprop



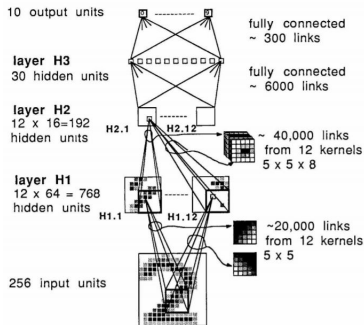
History (contd.)

- ① Rumelhart (1986) trained with backprop
- ② Showed that hidden units learn meaningful representations



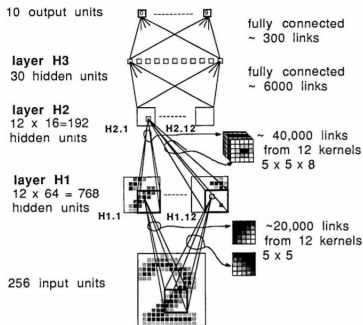
History (contd.)

- ① LeNet family (Lecun et al. 1989) is a “convent”



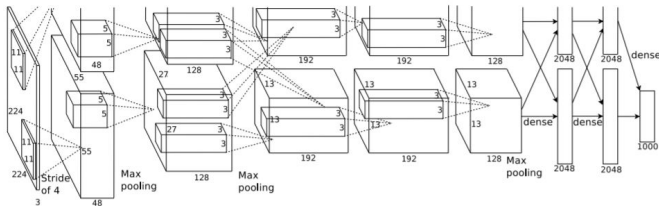
History (contd.)

- ① LeNet family (Lecun et al. 1989) is a “convent”
- ② Very similar to modern architectures



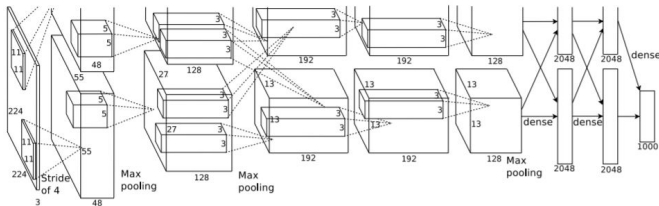
History (contd.)

1 AlexNet (2012)



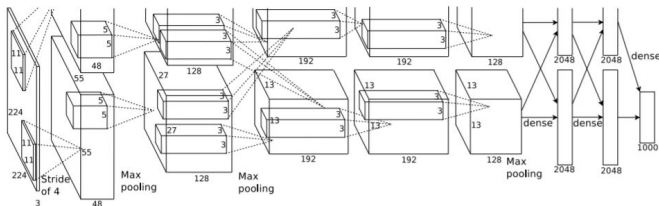
History (contd.)

- 1 AlexNet (2012)
- 2 Network similar to LeNet-5, but of far greater size



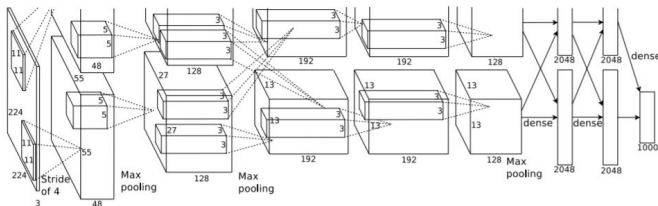
History (contd.)

- 1 AlexNet (2012)
- 2 Network similar to LeNet-5, but of far greater size
- 3 Implemented using GPUs



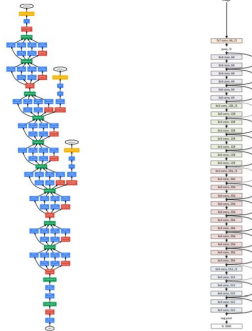
History (contd.)

- 1 AlexNet (2012)
- 2 Network similar to LeNet-5, but of far greater size
- 3 Implemented using GPUs
- 4 Could beat the SoTA image classification methods by a large margin



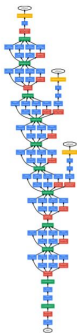
History (contd.)

- 1 AlexNet initiated a trend of more complex and bigger architectures



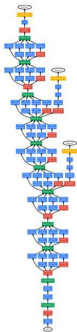
History (contd.)

- 1 AlexNet initiated a trend of more complex and bigger architectures
- 2 GoogLeNet (2015) contains “inception” modules



History (contd.)

- 1 AlexNet initiated a trend of more complex and bigger architectures
- 2 GoogLeNet (2015) contains “inception” modules
- 3 ResNet (2015) introduced “skip connections” that facilitate training deeper architectures



History (contd.)

① Transformers (2017) are attention-based architectures

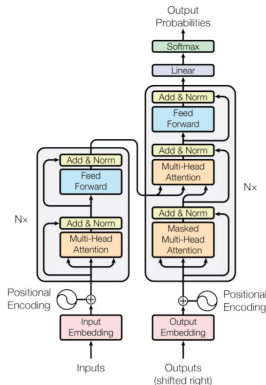


Figure credits: Vaswani et al., 2017

History (contd.)

- 1 Transformers (2017) are attention-based architectures
- 2 Very popular in NLP, and CV

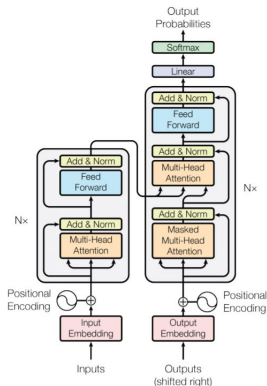


Figure credits: Vaswani et al., 2017

History (contd.)

- 1 Transformers (2017) are attention-based architectures
- 2 Very popular in NLP, and CV
- 3 Some of these models are extremely large - GPT-4 has 1.76 Trillion (estimated), Gemini 1.56 Trillion, PaLM-2 540B parameters

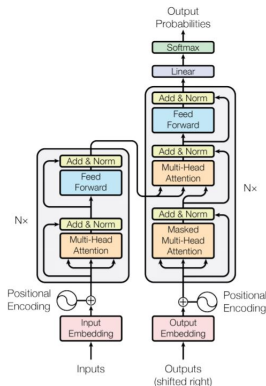


Figure credits: Vaswani et al., 2017

- ① Natural generalization to ANNs - Doesn't differ much from the 90s NNs

- ① Natural generalization to ANNs - Doesn't differ much from the 90s NNs
- ② Computational graph of tensor operations that take advantage of
 - Chain rule (back-propagation)
 - SGD
 - GPUs
 - Huge datasets
 - Convolutions, attention, Transformer, etc.

ILSVRC Error

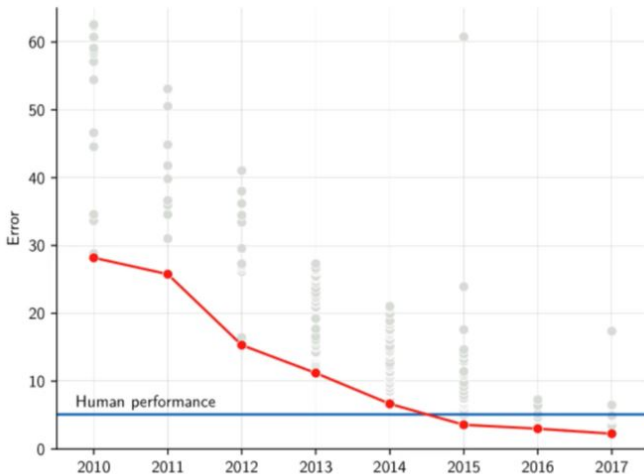


Figure credits: Gershgorn, 2017

LLM performance on the MMLU benchmark



Figure credits: W. Zi, L. El Asri, S. Prince

Deep Learning - practical perspective

- ① Doesn't require a deep mathematical grasp (!?)

Deep Learning - practical perspective

- ① Doesn't require a deep mathematical grasp (!?)
- ② Makes the design of large models a system/software development task

Deep Learning - practical perspective

- ① Doesn't require a deep mathematical grasp (!?)
- ② Makes the design of large models a system/software development task
- ③ Leverages modern hardware

Deep Learning - practical perspective

- ① Doesn't require a deep mathematical grasp (!?)
- ② Makes the design of large models a system/software development task
- ③ Leverages modern hardware
- ④ Doesn't seem to plateau with more data

Deep Learning - practical perspective

- ① Doesn't require a deep mathematical grasp (!?)
- ② Makes the design of large models a system/software development task
- ③ Leverages modern hardware
- ④ Doesn't seem to plateau with more data
- ⑤ Makes the trained models a commodity

Compute getting cheaper

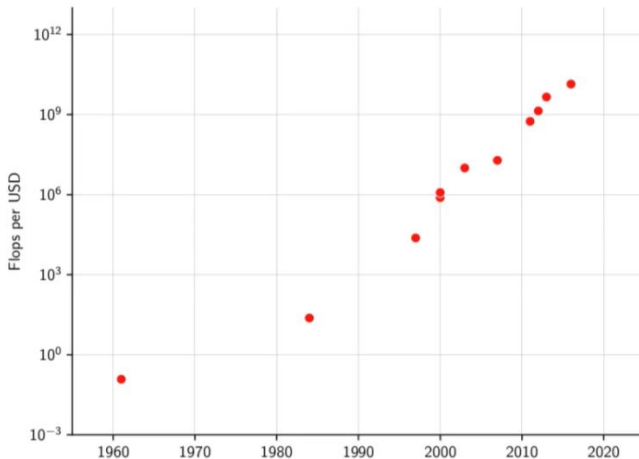


Figure Credits: Wikipedia

AlexNet to AlphaGo: 300000X increase in compute

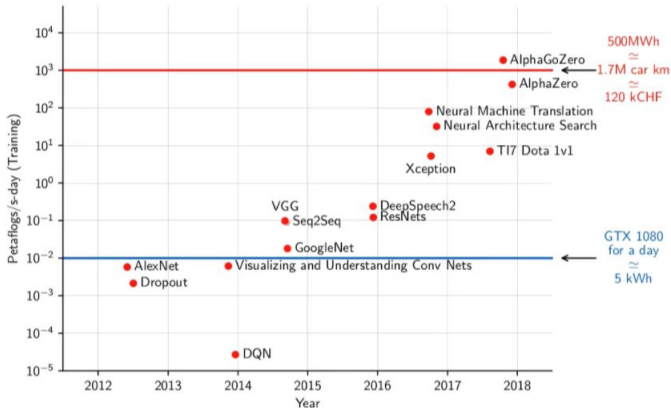


Figure Credits: Radford, 2018. 1 petaflop/s-day \approx 100 GTX 1080 GPUs for a day, \approx 500kwh

Storage getting cheaper

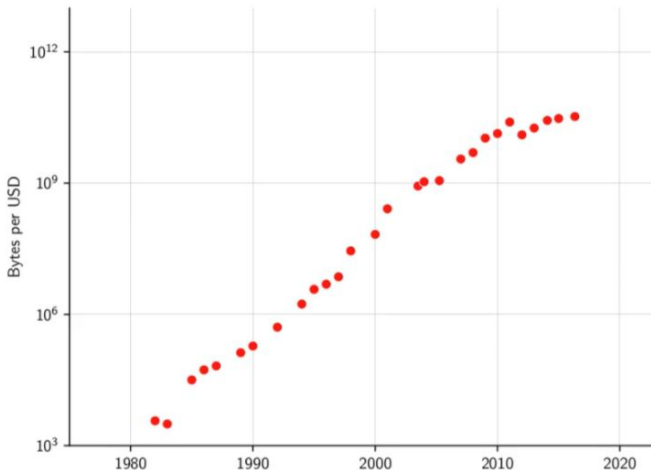


Figure Credits: John C Mccallum

LLM compute

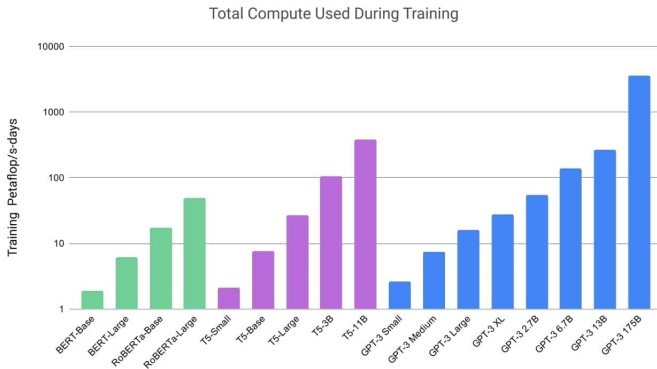


Figure Credits: [NVIDIA blog](#)

Datasets

Data-set		Year	Nb. images	Size
MNIST	(classification)	1998	60K	12Mb
Caltech 101	(classification)	2003	9.1K	130Mb
Caltech 256	(classification)	2007	30K	1.2Gb
CIFAR10	(classification)	2009	60K	160Mb
ImageNet	(classification)	2012	1.2M	150Gb
MS-COCO	(segmentation)	2015	200K	32Gb
Cityscape	(segmentation)	2016	25K	60Gb

Data-set		Year	Size
SST2	(sentiment analysis)	2013	20Mb
WMT-18	(translation)	2018	7Gb
OSCAR	(language model)	2020	6Tb

Figure Credits: François Fleuret

- GPT-3 uses 45TB of text data for training

- GPT-3 uses 45TB of text data for training
- LAION dataset consists of 5.85 billion image-caption pairs

Implementation

	Language(s)	License	Main backer
PyTorch	Python, C++	BSD	Facebook
TensorFlow	Python, C++	Apache	Google
JAX	Python	Apache	Google
MXNet	Python, C++, R, Scala	Apache	Amazon
CNTK	Python, C++	MIT	Microsoft
Torch	Lua	BSD	Facebook
Theano	Python	BSD	U. of Montreal
Caffe	C++	BSD 2 clauses	U. of CA, Berkeley

Figure Credits: François Fleuret

- Please visit **lectures tab in the course website** for the full list of references
- Please share your comments/suggestions/any errors (technical or otherwise) with the instructor (**krmopuri@ai.iith.ac.in**)
- Thank You!

Next class

- Artificial Neuron Models