Machine Learning for Jet Physics

RTG Student Lecture

Peter Sorrenson

Contents

Lecture 1

ML history

- Main idea of ML: learn statistical patterns from data in order to perform tasks without explicit programming
  - Driving force: continually dropping cost of computation and data (*Two Centuries of Productivity Growth in Computing*, Nordhaus 2007)

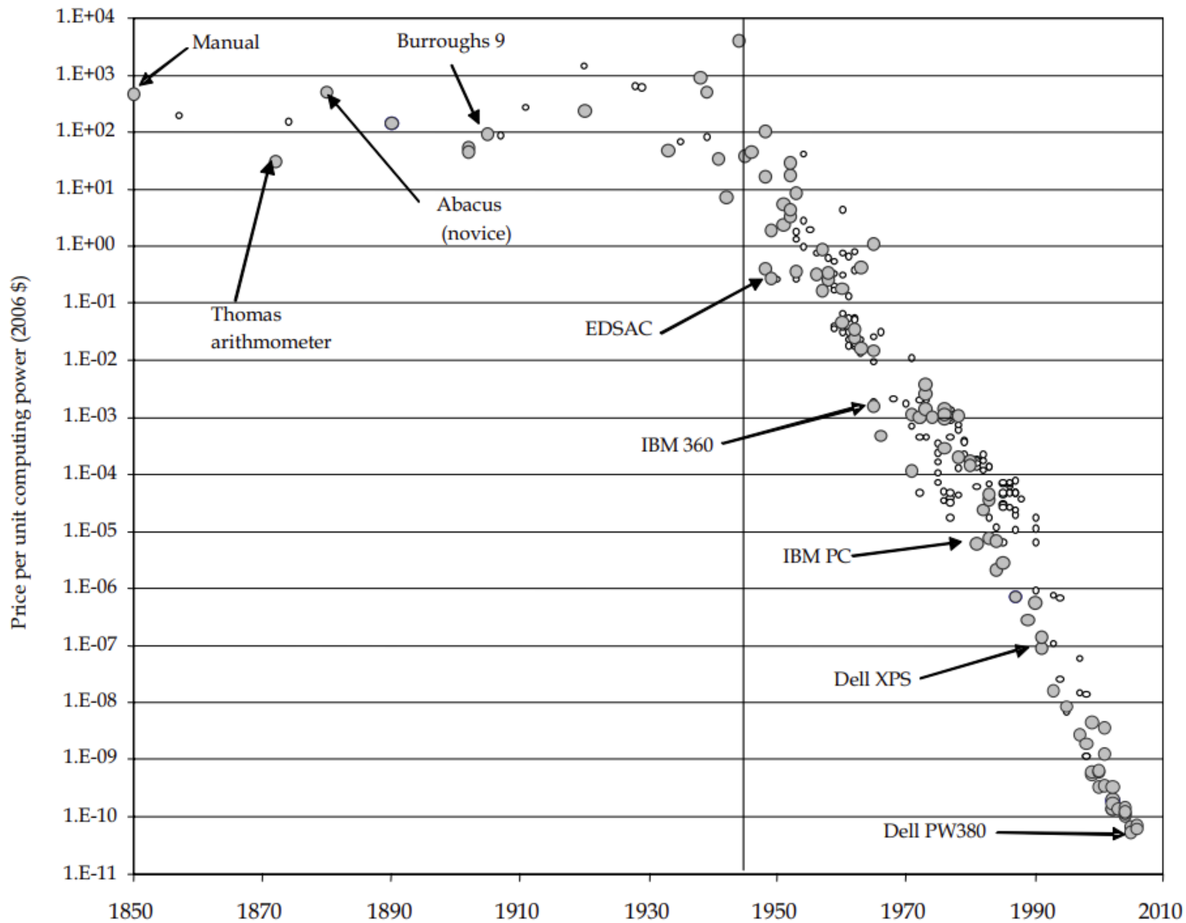
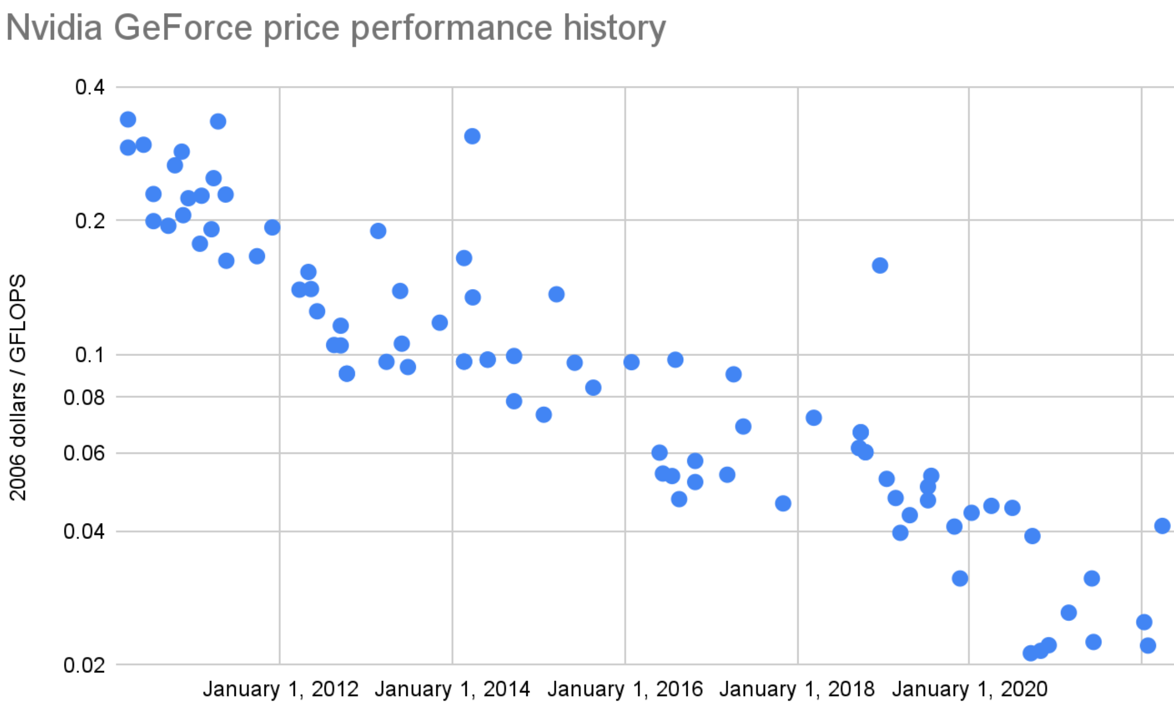
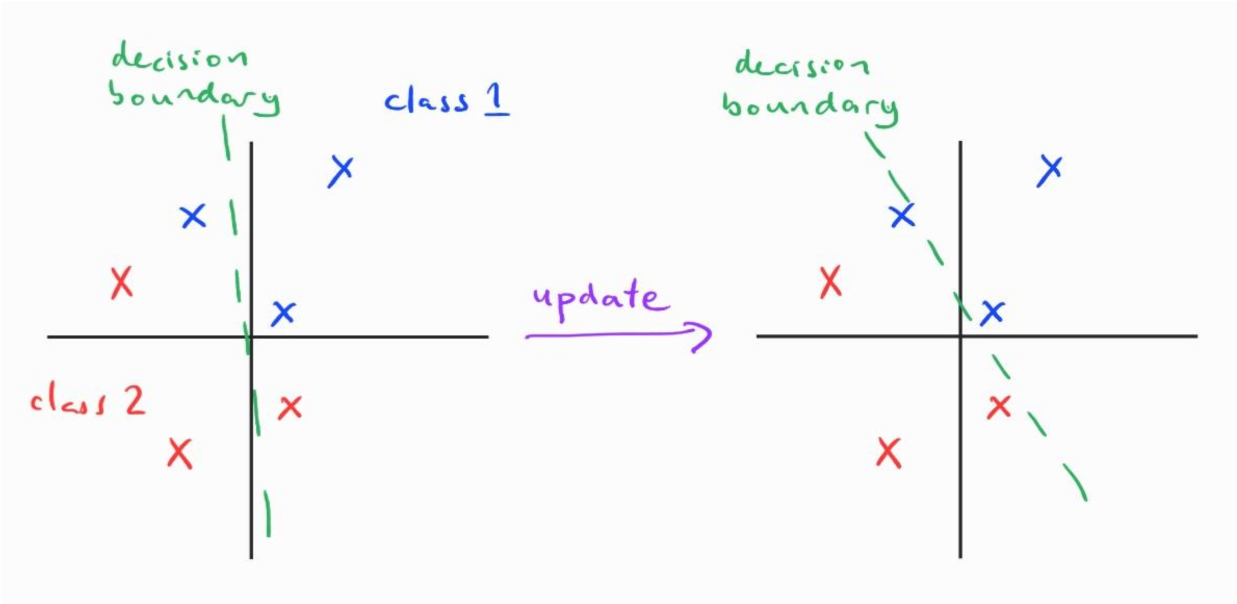


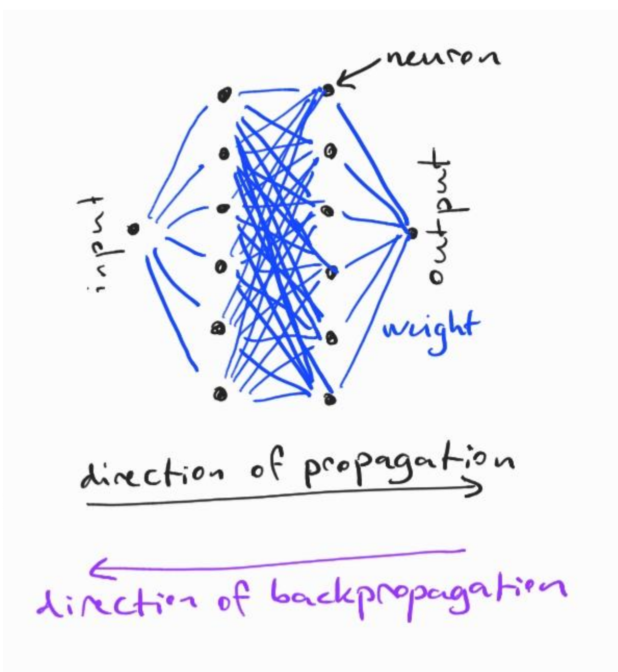
FIGURE 3
THE PROGRESS OF COMPUTING MEASURED IN COST PER COMPUTATION PER SECOND DEFLATED BY THE PRICE INDEX FOR GDP IN 2006 PRICES

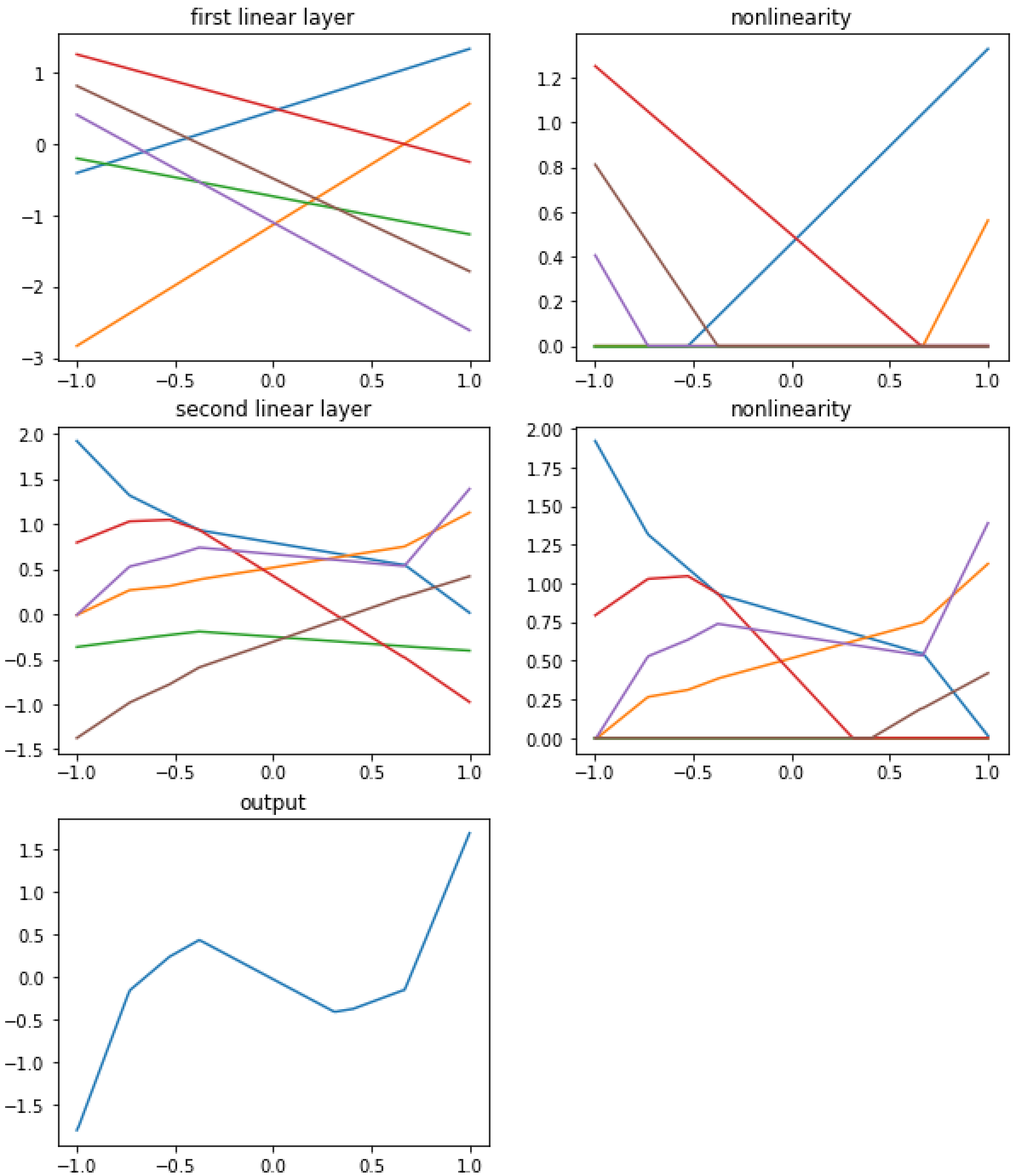


    - Especially good at hard-to-define tasks that work in high-dimensional spaces, e.g. classifying an image
    - Software 2.0: rather than programming a function (software 1.0), provide data that defines the input -> output pair
        - It turns out that for many interesting problems it's significantly easier to collect the data than to define the program
    - “Every time I fire a linguist, the performance of our speech recognition system goes up” (Fred Jelinek from 1985)
- Invention of transistor (1947) and Shannon’s theory of information (1948)
    - Events which kicked off the information age
- Perceptron (1957)
    - Linear model which is updated iteratively to find a decision boundary
    - Sketch of perceptron update:

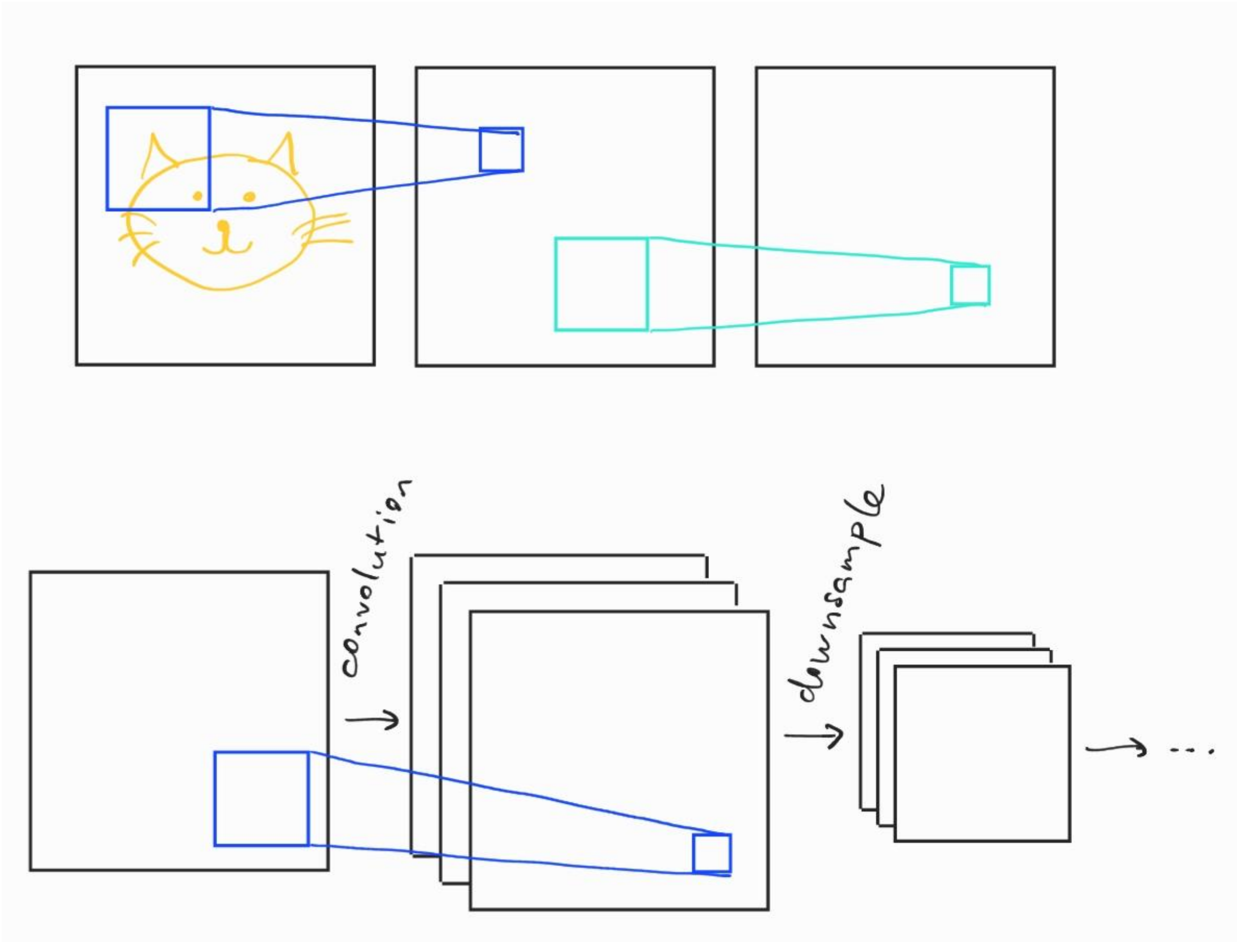


- Backpropagation (first described 1970, rediscovered 1986)
    - Allows training of multiple-layer neural networks
    - Example of three-layer network: 1D input to 1D output with 2 hidden layers of 6 neurons each

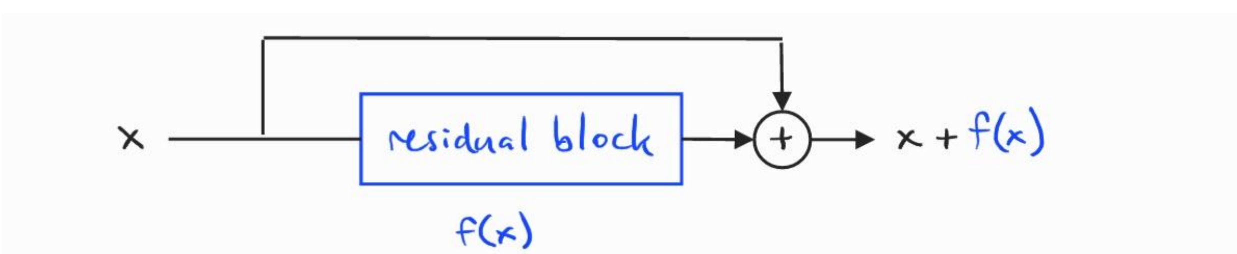




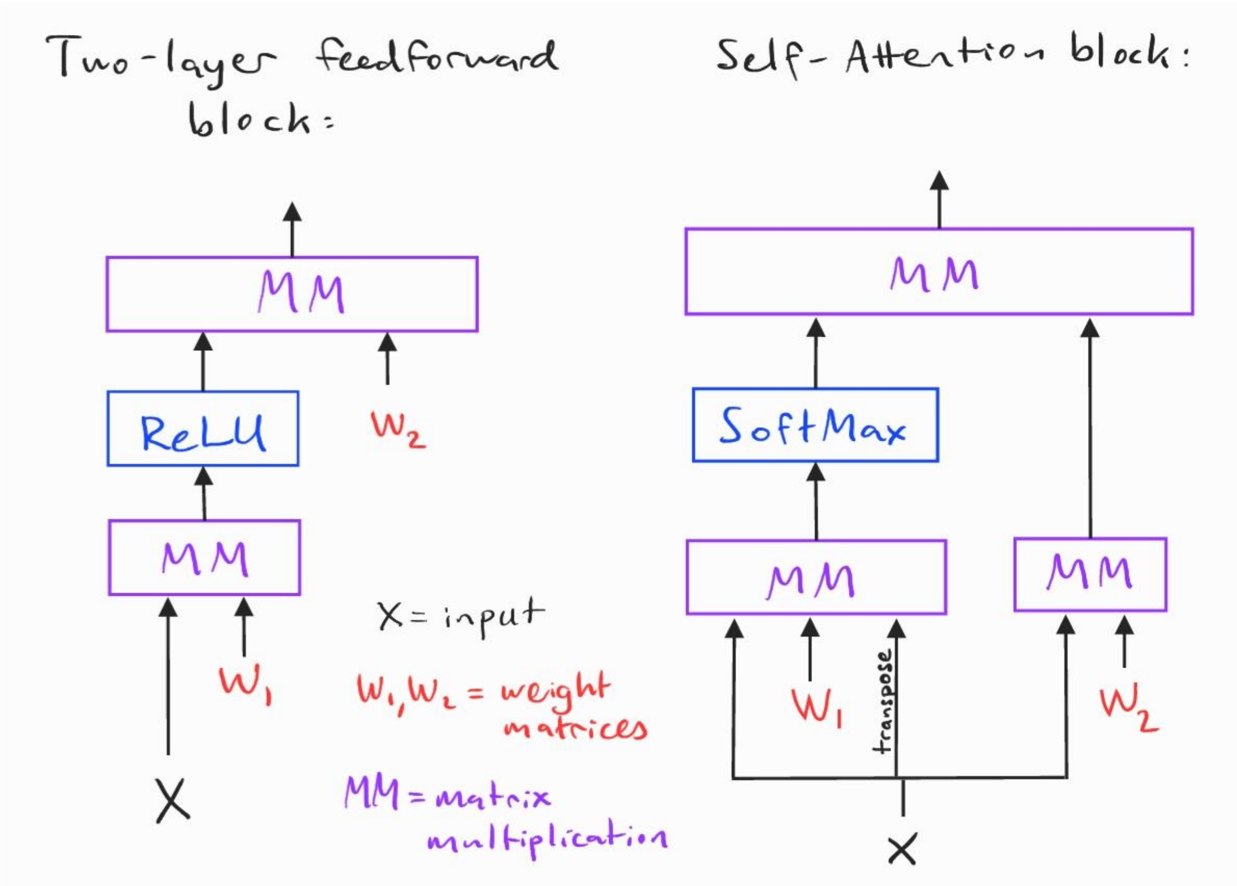
- Convolutional neural net (LeNet 1989)
  - Efficient image processing
  - Exploits translational symmetry of images
  - Sketch of stacked convolutional layers:



- Deep Blue beats Kasparov (1997)
- Development of GPUs (2000s)
- Deep neural nets (2010s)
  - ImageNet (2009)
    - Huge labeled dataset which kicked off deep learning
  - AlexNet (2012)
    - The model which made it clear that neural nets are the most powerful models around (at least for computer vision)
    - Used custom CUDA kernels to speed up convolutional nets
  - ResNet (2015)
    - Allowed very deep models, template for what followed
    - Sketch of residual blocks:

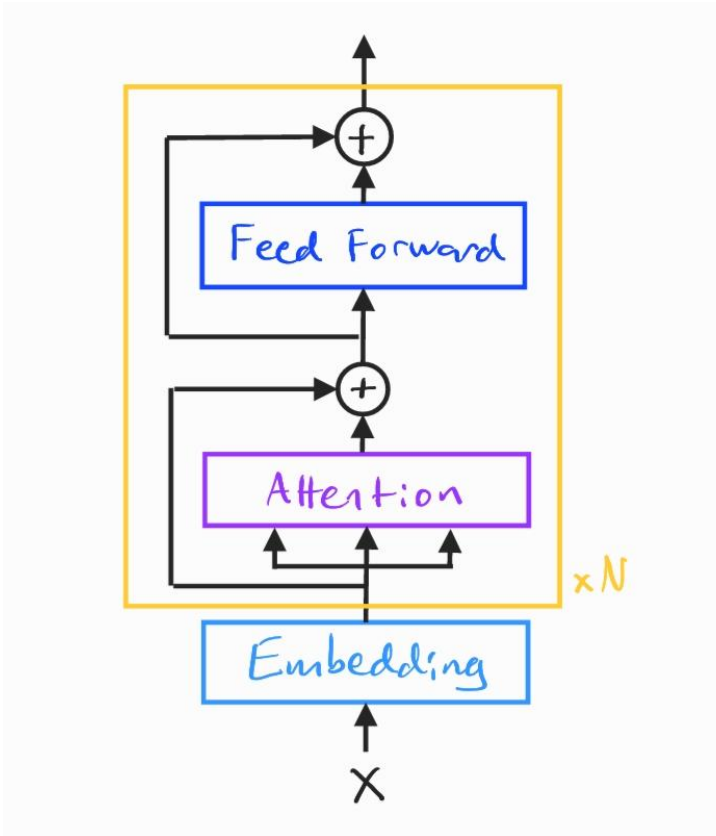


- AlphaGo (2016)
    - Maturation of reinforcement learning through self-play in discrete perfect-information games
    - AlphaZero a year later
- Transformer (2017)
    - General purpose model which can be applied to almost any type of data
    - Extensive use of attention mechanisms:



    - Attention is powerful because it allows for the expressive modeling of interactions within a sequence or set of data. Most data can be expressed as a sequence or a set (e.g. an image is a sequence of small patches)

- Sketch of transformer structure:



- GPT-3 (2020)
  - Large language model with surprisingly good fluency in natural language and ability to generalize and perform logical tasks
- AlphaFold 2 (2021)
  - Protein folding prediction to unprecedented accuracy
- The basics remain the same: sequential computations based on large matrix multiplications, trained via gradient descent on lots of data
- Interactive tool: see effect of NN layers and play around at https://playground.tensorflow.org/