Discussion

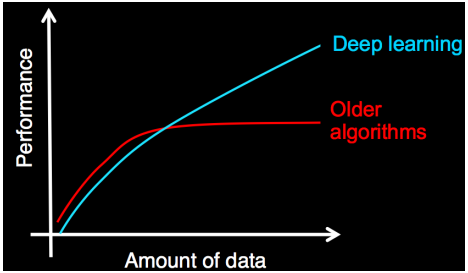
He He

CDS, NYU

April 20, 2021

Advantages of Neural Networks

Neural Networks Benefit from Big Data



- Empirical observation: performance of deep neural networks has not plateaued with increasing amount of data.
- Higher data throughput compared to other nonlinear methods.
- Recent trends in system for ML: how to run (training/inference) large neural networks efficiently.

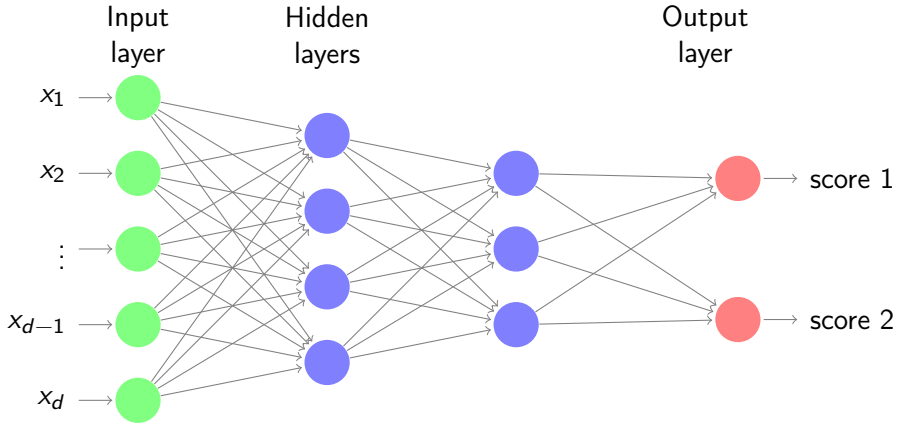
From Andrew Ng's CS229 Deep Learning slides (http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf)

Versatile architectures

- Easy to incorporate inductive bias for different tasks.
- Examples:
  - Translation invariance in image recognition—convolution.
  - Dependence on past observations in time series—recurrence.
  - Alignment between input and (structured) output—attention.
- Many building blocks can be composed together.

Representation sharing

- "Classifiers" are task-specific but representation/features can be shared.



Multitask Learning and transfer learning

Multitask learning:

- Learn related tasks together, e.g.,
  - Object classification: cat or dog?
  - Object localization: location of the objects?
- Basic features (e.g., edges, texture) can be shared by tasks.
  - Different output layers for each task; the rest is shared.
  - Objective function combines losses from both predictions, e.g. by averaging.

Transfer learning:

- Self-supervised **pre-training** to learn generic features.
  - General idea: denoising, i.e. perturbed input $\rightarrow$ original input.
- On downstream tasks: **fine-tune** pre-trained models (reuse representation).

Summary

- Powerful use of features: representation learning
- Fast and scalable with data given the right system support.
- Hard to train: non-convex optimization
  - Easier in practice with released code and libraries.
- Gap exists between theory and practice: when and why does it work?