Motivation for Deep Learning

Based on chapter5 and chapter12

Linara Adilova

Why Do We Need Deep Learning?

- Before introduction of large neural networks and set of tools connected to DL, machine learning algorithms were **failing** in solving important problems in AI:
  - recognizing speech
  - recognizing objects
  - etc.
- The development of DL was motivated in part by the failure of traditional algorithms to **generalize** well on such AI tasks

Larger Requirements to Models

- XOR problem cannot be solved with a network without hidden layers
  - One hidden layer with nonlinearity solves the problem
- MNIST dataset for zip codes recognition
  - Convolutional network
- Imagenet dataset with thousands of classes of images
  - ResNet, etc.

Problem: Curse of Dimensionality

- Many machine learning problems become exceedingly difficult when the number of dimensions in the data is high - **curse of dimensionality**
- Of particular concern is that the number of possible **distinct configurations of a set of variables** increases exponentially as the number of variables increases

Problem: Curse of Dimensionality

- Simple assumption that it will be same value as the **neighboring examples** becomes senseless
  - might be no examples in “neighborhood”
  - possible “neighborhood” definition is too vague

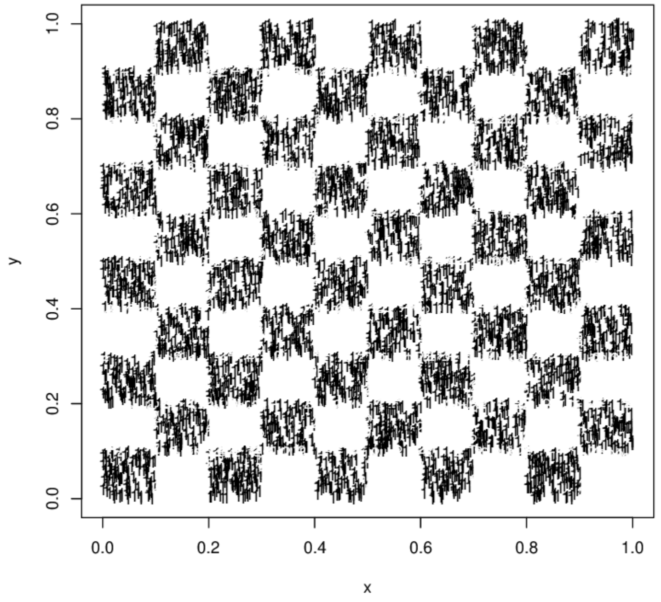
- DL models can consider random and large configurations automatically

Problem: Smoothness Prior

- In order to **generalize well**, machine learning algorithms are guided by prior beliefs about what kind of function they should learn
  - directly influence the function itself
  - indirectly acting on the parameters via their effect on the function
  - choosing algorithms that are biased toward choosing some class of functions over another

Problem: Smoothness Prior

- The **smoothness assumption** work well as long as there are enough examples to observe high points on most peaks and low points on most valleys of the true underlying function



- In DL we assume that the data was generated by the **composition of factors** or features, potentially at multiple levels in a hierarchy

Problem: Manifolds

- In machine learning term **manifold** is used to designate a set of points that are similar to each other inside of a more diverse space
  - variations in data happen only when we move from one manifold to another
- **Manifold hypothesis** suggests that real world data is defined by manifolds
  - the probability distribution over images, text strings, and sounds that occur in real life is highly concentrated
    - face images in overall distribution
  - informal neighborhoods and transformations between manifolds
    - in the case of images gradually move or rotate objects in the image
- DL aims at learning the manifold structure of the data

Large Scale Implementation

- DL is based on the philosophy of **connectionism**
  - an individual biological neuron is not intelligent, a large population of these neurons acting together can exhibit intelligent behavior
- Important to emphasize that the number of neurons must be **LARGE**
- Artificial neural network sizes have **grown exponentially** for the past three decades, yet they are only as large as the nervous systems of insects
- Multiple approaches to optimise and parallelise

Application: Computer Vision

- Mostly is used for **object recognition** some form:
  - which object is present in an image
  - annotating an image with bounding boxes around each object
  - transcribing a sequence of symbols from an image
  - labeling each pixel in an image with the identity of the object it belongs to
- Also **image synthesis**, e.g., for image restoration

Application: Speech Recognition

- The task of speech recognition is **mapping an acoustic signal** containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker
- 2009: speech researchers applied a form of DL based on **unsupervised learning**
  - restricted Boltzmann machines
- With **larger labeled datasets** the unsupervised pretraining phase was not bringing any significant improvement
- Main directions
  - **Convolutional networks** that replicate weights across time and frequency
  - End-to-end deep learning speech recognition systems, e.g., a **deep LSTM RNN**

Application: Natural Language Processing

- An efficient model of natural language uses techniques that are specialised for processing **sequential data**
- A **language model** defines a probability distribution over sequences of tokens (a word, a character, or a byte) in a natural language
    - **Neural language models** are a class of language model using a distributed representation of words (**word embeddings**)
- **Scaling problems solutions**
    - short vocabulary,
    - hierarchical softmax,
    - importance sampling

Ethics of Applying Powerful ML

- Danger of overusing autonomous decisions
  - No responsibility
  - No control
- Ethics of using data
- “Face recognition and the ethics of AI”

(https://www.ben-evans.com/benedictevans/2019/9/6/face-recognition)