CS F425: Deep Learning

Tirtharaj Dash

Dept. of CS & IS and APPCAIR
BITS Pilani, Goa Campus

August 21, 2021

Course Staff

Instructor-in-Charge:

Tirtharaj Dash

Teaching Assistants:

Atharv Sonwane

Anmol Agarwal

Ameya Laad

Akhilesh Adithya

The Course I

DL success can be attributed to:

1. Availability of large amount of data
2. Extensive use of powerful graphics processors
3. Availability of software libraries to facilitate deep network implementations
4. Significant involvement of industrial research labs in deep learning research.

The Course II

Course Plan:

1. Preliminaries
2. Multilayered Neural Nets
3. Neural Nets for Computer Vision
4. Neural Nets for Sequence Learning
5. Neural Nets for Representation Learning
6. Neural Nets for Generative Modeling

The Course III

Course Websites:

- Classroom: Google Classroom
- Lab: CS-F425_Deep-Learning (GitHub)

The Course IV

Evaluative Components:

Component	Mark	Type
Midsem Exam	30	Open Book (Online)
Lab Assignment 1	10	-NA-
Lab Assignment 2	10	-NA-
Major Project	20	-NA-
Comprehensive Exam	30	Open Book (Online)

The Course V

I will NOT teach:

- Machine Learning basics
- Optimisation basics

Any questions?

A brief look back ... I

Inspiration of Deep Learning and its history

- Loosely: Brain :: Neural Nets [like, Bird :: Aeroplane; in fact, Deep Learning has nothing to do with human brain.]
- Historical names: Cybernetics (1940s–1960s), Connectionist Models (1980s–1990s), Deep Learning (2006–)
- Neural Network (NN, Neural Net) is the term used to refer to such an architecture.

A brief look back ... II

- Started: McCulloch and Pitts Model of neuron (1943)
  - Idea: Neurons are threshold units (on/off states)
  - Purpose: Build Boolean circuit by connecting neurons
  - Outcome: Perform logical inference
  - How: (1) Neurons compute weighted sum of inputs; (2) Compare the sum to its threshold; (3) Neuron is turned 'on' if the sum is above the threshold; 'off' otherwise
  - A simplified view of how a neural network works

A brief look back ... III

- Donald Hebb: Hebb's rule or Hebbian Learning (1947)
  - Idea: Neurons in the brain learn by modifying the strength of the connections between neurons
  - How: If two neurons fire together, the connection linked between them increases; decreases otherwise
  - Also called hyper learning

A brief look back ... IV

- Norbert Wiener: Proposal for cybernetics (1948)
  - Idea: having systems with sensors and actuators, you have a feedback loop and a self-regulatory system
  - Result: The rules of the feedback mechanism of a car all come from this work.

A brief look back ... V

- Frank Rosenblatt: Perceptron (1957)
  - Weight modification in a simple neural net
  - This was a big breakthrough in the field

A brief look back ... VI

- Towards late 1960s, the field started to die off. Reasons:
  - The researchers used neurons that were binary (not differentiable)
  - There was no idea of continuous neurons (or, activation functions)
  - Backpropagation requires continuous activation function
  - Before 1980: the multiplication of two floating-point numbers were extremely slow

A brief look back ... VII

- Restarted again: 1985 with emergence of backpropagation

A brief look back ... VIII

- 1995: the field died again and the machine learning community abandoned the idea of neural nets

A brief look back ... IX

- 2006-2010:
  - Huge performance improvement in speech recognition tasks using neural nets
  - Wide deployment in the commercial field

A brief look back ... X

- 2013: Computer Vision switched to neural nets
- 2016: Natural Language Processing switched to neural nets

- ..., and the rest is history!

A brief look back ... XI

Supervised Learning

- Majority of deep learning applications use supervised learning.
- Steps:
  - Collect a bunch of pairs of inputs and outputs
  - Inputs are feed into a machine to learn the correct output
  - When the output is correct, don't do anything
  - If the output is wrong, tweak the parameter of the machine and correct the output toward the one you want.
  - Change direction and amount of update requires gradient computation and backprogation

A brief look back ... XII

Pattern Recognition (before emergence of DL):

- Data $\rightarrow$ Feature Extraction $\rightarrow$ Trainable Classifier
- Issue: The feature extractor was designed by hand.

A brief look back ... XIII

Pattern Recognition (in DL era):

- Sequence of modules (each module is a feature extractor)
- Each module has tunable parameters (and nonlinearity)
- Modules are stacked one after another (a "deep" stack)

Deep Learning: Components

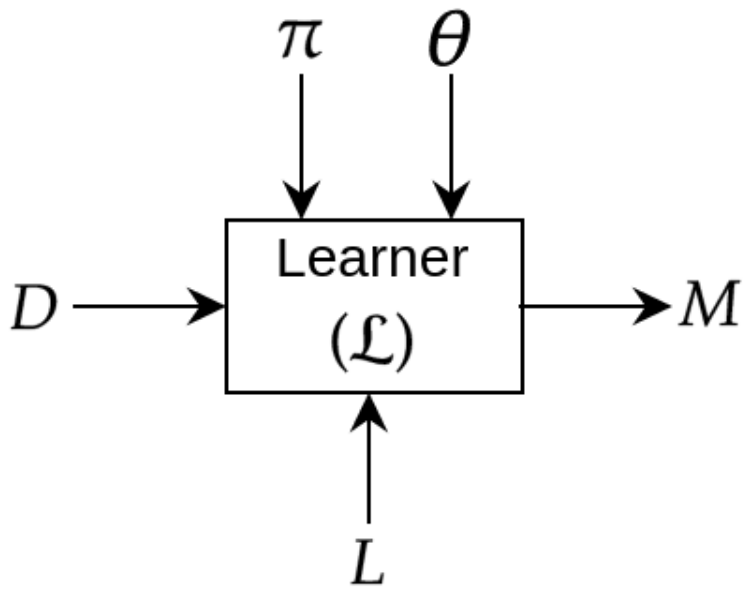
D: Dataset

M: Model

π: Structure of the model

$\boldsymbol{\theta}$: Parameters of the model

L: A loss function



(T.Dash et al.: https://arxiv.org/abs/2107.10295)

I expect you to study the basics of the above components yourself. See the next slide for the relevant references to these.

Self-study

Study the following to get comfortable with various components discussed in the previous slide.

- Machine Learning and Optimisation basics: Ch.5 of Textbook-2
- Introduction and Linear Neural Networks: Ch.1–3 of Textbook-1