Lecture1 (History, Evolution and revolution of DL) 21/12/2020

Inspiration of DL: The Brain

-> McCulloch & Pitts (1943): networks of binary neuron can do logic

-> Donald Hebb (1947): Hebbian synaptic plasticity

-> Norbert Wiener (1948): cybernatics, optimal filter, feedback, autopoises, auto-organization

-> Frank Rosenblatt (1957): Perceptron

-> Hubel & Wiesel (1960s): visual cortex achitecture

stopped and again started in 1980s and died in 1995 again and again rose up in 2010s with speech recognition.

2012-2013 -> rise of Computer Vision

2016 -> rise of NLP

Supervise Learning: Training a machine by showing examples instead of programming it.

Works well for:

Speech -> words
Image -> categories
Portrait -> name
Photo -> caption
Text -> topic

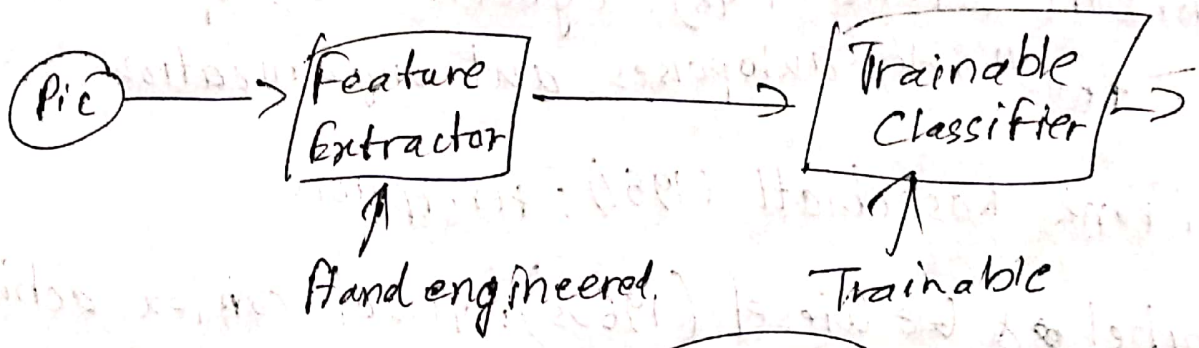
Supervised Learning goes back to the Perceptron and Adaline

The McCulloch-Pitts Binary Neuron

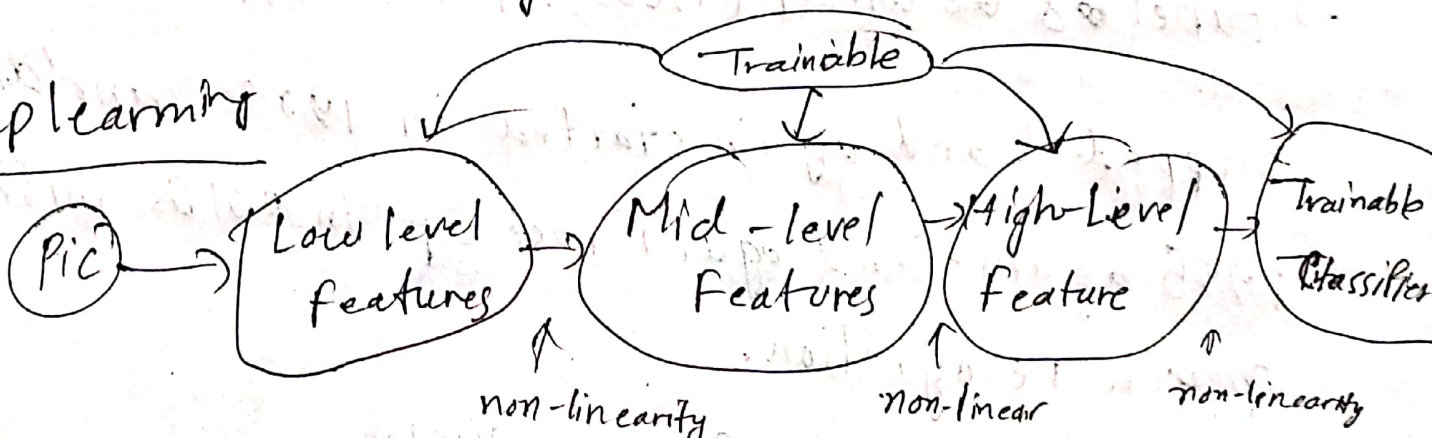
→ Perceptron: Weights are motorized potentiometers

→ Adaline: Weights are electrochemical memristors

Traditional Machine Learning



Deep learning



MLP (Multi-layer Perceptron)

$$ReLU(x) = \max(x, 0)$$

Supervised Machine learning = function Optimization

Stochastic gradient descent

$$W_i \leftarrow W_i - \eta \frac{\partial L(W, X)}{\partial W_i}$$

Computing Gradients by Back-Propagation:

$\frac{dC}{dWn}$ ⟶

Traditional NN has usability problems and many others dealing with images.

- Hubel and Wiesel's Model of the Architecture of the Visual Cort...
  - simple cells detect local features [1962]
  - complex cells "pool" the outputs of simple cells within a space
    → They found out how things work in brains for recognition

[Fukushima 1982]
[LeCun 1989, 1998]

ConvNets can recognize multiple objects

- All layers are convolutional
- Networks performs simultaneous segmentation and recognition

Face and Pedestrian Detection with ConvNets (1993-2015)
Training a Robot to Drive itself in Nature [R Hadsell 2009]
Semantic Segmentation with ConvNets [Farabet 2012]

1986-1996] ⇒ Special NN chips

Deep learning Revolution

Speech Recognition 2010

Image Recognition 2013

NLP 2015

Deep ConvNets for Object Recognition (on GPU)

AlexNet [2013] OverFeat [2013]

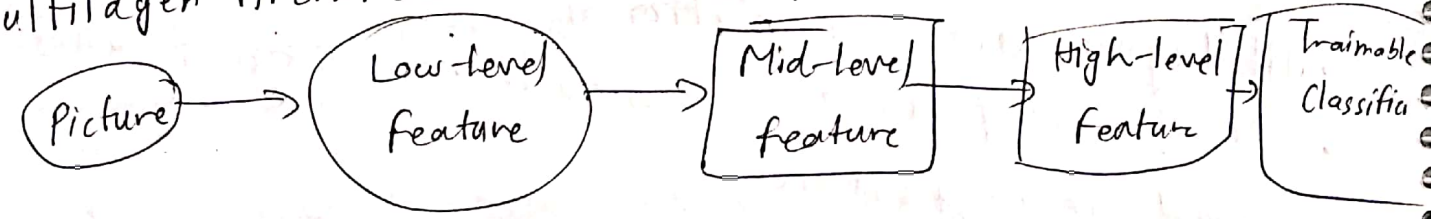
VGG [2013]

GoogleNet [2013]

ResNet [2015]

DenseNet [2017]

Multilayer Architectures == Compositional structure of Data



MSRA - 2015 Mask R-CNN 2017

Mask R-CNN 2017

Mask-RCNN on COCO dataset

→ Individual objects are segmented

3D ConvNet for Medical Image Analysis

Deep learning enables

→ Safer cars, autonomous cars

→ Better medical image analysis

→ Personalized medicine

→ Adequate language translation

→ Useful but stupid chatbots

→ Info search, retrieval, filtering

→ other

It can't yet

Machines with common sense

Intelligent personal assistants

Smart chatbots

Household robots

Agile and dexterous robots

Artificial General Intelligence

Deep learning = Learning Representations/Features

Hierarchical representation

→ Hierarchy of representations with increasing level of abstraction

→ Each stage is a kind of trainable feature transform

SVM is nothing but two layer neural nets and the first layer is trained in an unsupervised way

→ Deep machines are more efficient for representing certain classes of functions

Why would DL be more efficient? [Bengio, Lecun 2007]

→ less params needed to get good result

What are Good features?

Discovering the Hidden structure in High Dimensional Data: The Manifold Hypothesis

The Manifold Hypothesis

- Natural data lives in a low-dimensional [non-linear] manifold
- Because variables in natural data are mutually dependent

Disentangling factors of variation

- The ideal disentangling feature Extractor
- PCA can find the representation if they are linear