Machine Learning (CS60050)
Instructor: Aritra Hazra

Scribes By:
Km Simran Jaiswal (20CS60R58)
Lecture Date: 7^{th} April 2021 , Wednesday

1 Introduction

Given genesis of Artificial Intelligence (AI) machine can think by using planning and heuristics algorithm i.e. machine can enact thinking of human beings. Then,machine can learn using the concept of machine learning.
In the era of big data machine can emulate vision perspective or sequence of operating perspective of human beings. This dimension of learning gave rise to deep learning.The concept of deep learning is driven from artificial neural network. [1]

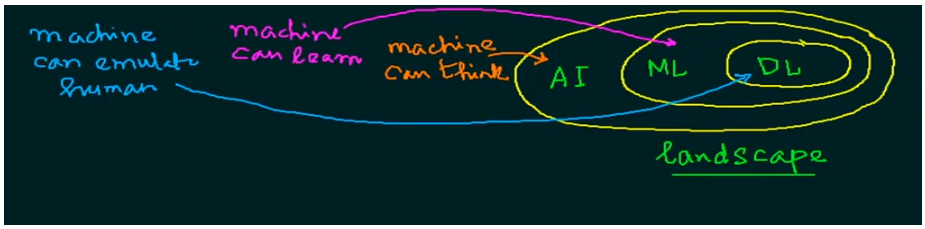

Figure 1: Landscape of AI,ML and DL

2 Deep Learning

The concept of deep learning is driven from artificial neural network. In term of neural network let x_1 and x_2 are input variable.Bias is w_0 with weight 1.The feed-forward neural network is shown in Fig(2).
The neural network classifier classifies the input with respect to different boundaries it can draw.By different stacking of layer in neural network we can draw boundary and data points get segregated.The boundaries to classify data points using neural network is shown in Fig(2). In feed forward neural network every

[1]This scribe is based on lecture taught by Prof. Aritra Hazra on 07 April 2021 in Machine Learning(CS60050) course.

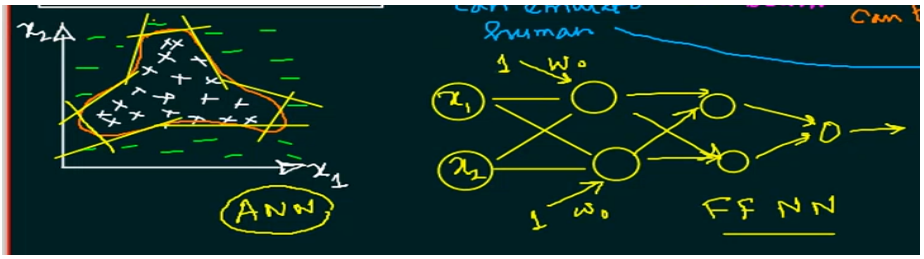

Figure 2: Feed forward neural network

layer takes input from every other previous layer. In fig(3) we can see that layer l_{i-1} feeds input to layer l_i.The two neuron in layer l_i will compute $\theta(\sum w_i * x_i)$ and $\theta(\sum w_i * x_i)$.In this model of neural network spatial information is lost.

2.1 Feature Extraction

In feature extraction we take decision on basis of composition of features we get. Suppose there is an image.It is basically pixel value at each location.Let say there is 4x4 image of pixel value $x_1 \dots x_{16}$ as seen in fig(3). Now when we feed this image into network we focus on positional aspect of pixels not the full image. We extract high level feature from them.
The "depth" of neural network is number of layers in network.
The "breadth" of each layer is the number of neurons in each layer.It is for the number of features we are distinguishing and trying to reason about.In visionary system,given analogy to human it is how many different zones we are looking and trying to make decisions. As we move across depth we compose these features. Feature extraction can be hierarchical. In each level we can extract features

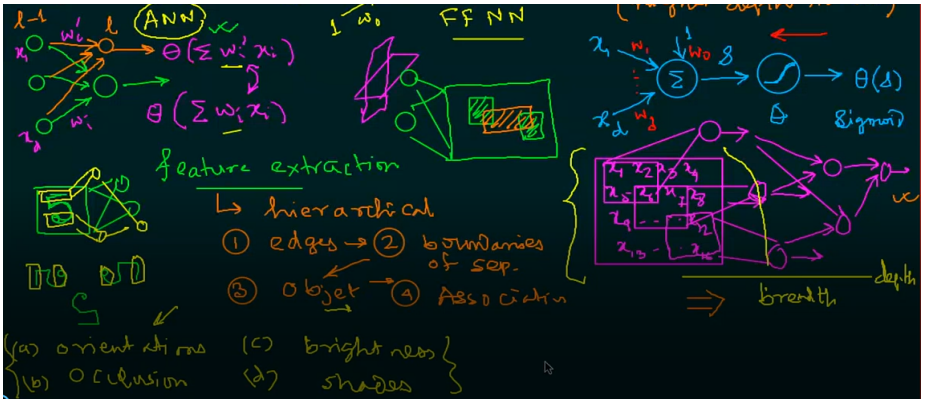

Figure 3: Feature Extraction

such as:

1. At level 1 we extract edge of image.
2. At level 2 we extract boundaries of separating edges.
3. At level 3 we compose them and separate object.
4. At level 4 we see association of object.

By feature extraction we can tackle following properties of data:

1. Orientation
2. Brightness
3. Occulsion
4. Shades

2.2 Convolutional Neural Network

In convolutional neural network we extract features from an image in order to understand the objects and then by composing this features we can extract them.
Suppose 'X' can be written in two ways ('X' and 'χ').The images with pixels are shown in fig(4). The pixel which are black (shaded) is +1 and pixel which are white are marked as -1. Now we do feature-wise comparison (enclosed in yellow box in 'X' and orange box in 'χ') in fig(4).

To build a neural network we need to decide which extracted unit to take

Figure 4: Pixel Image of 'X' and 'χ'

forward. The operation required to build neural network are:

1. **Convolution operation :** Suppose there is 3x3 kernel matrix: $\begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$
   If we keep on scanning 3x3 zones overlapping wise and perform matrix multiplication of each zone with kernel matrix we get the maximum value as 9(For zone highlighted pink in color in fig(5)).No other zone will get

this value. This will give us centre of 'X'. Hence,on convoluting the zone the zone which give us maximum value will be centre of 'X'.
Given image matrix of 5x5 we convolute it with kernel matrix of 3x3 and get the extracted features.

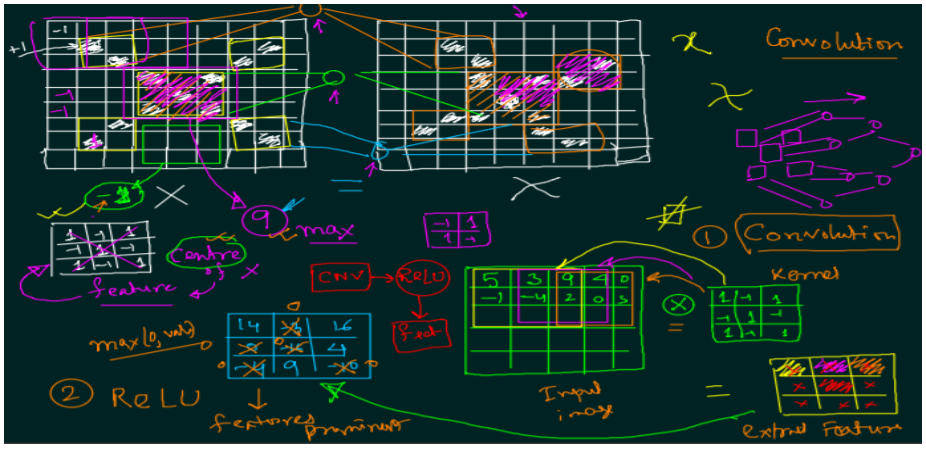

Figure 5: Convolution Operation and ReLU

2. **ReLU (Rectified Linear Unit):** The output of convolution operation i.e. feature extracted will have some values which will be significantly higher and there will be some values which will be significantly lower. We eliminate this insignificant values and replace it with zero.

$$val = max(0, val)$$

The ReLU makes feature prominent.

Suppose we are trying to recognize the car. The CNN to identify car is shown in Fig(6).

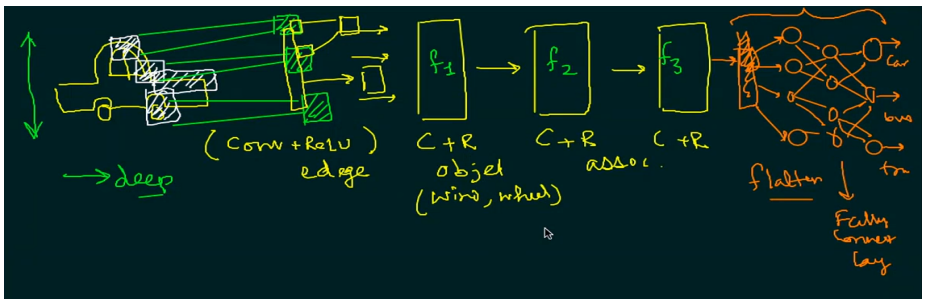

Figure 6: CNN to identify Car

How to select kernel?

At each layer of CNN we train weights. Along with that we need to train value of kernel.This thing can be applied only if we have huge amount of training examples.The choice of kernel should dictate whether we want to smoothen the image or sharpen the image.

Demerits of CNN :

- Training time is high.
- Requires high amount of training examples.
- Lost of Explainability aspect.

Applications of CNN: Breast Cancer Detection.

2.3 Transfer Learning:

We have CNN model (M_1)that recognizes object.Can we use this M_1 to recognize X-Rays. To perform Transfer Learning two steps are required:

Step 1: Pre-training

Step 2: Fine Tuning

Let $\hat{y}$ be predicted value from model M_1. To predict X-rays we need to change model M_1. We add new layer in model M_1. This new layer has to learn new weights w_{new}.This model to recognize X-rays be M_2.

One thing to be noted for transfer learning is that number of training examples

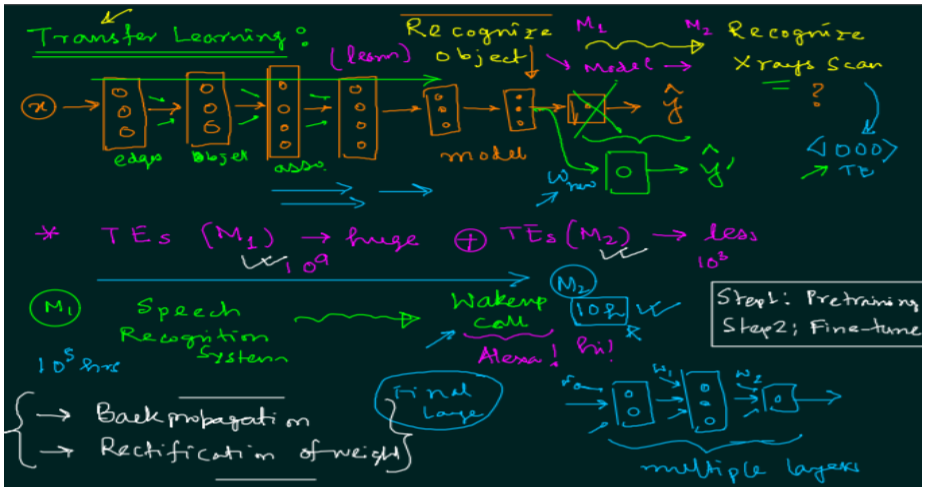

Figure 7: Tranfer Learning

to build M_1 must be huge (10^9) as compared to M_2 (10^3).

In speech recogniztion system (10^5 hours of training data) we need to transfer this knowledge to wakeup call alarm.(10 hours of training data).

3 Conclusion

In this lecture we studied about introduction to deep learning and convolutional neural network.