Neural Machine Translation

Seminar Report-NLP

Arul, Ehsan

Rheinische Friedrich-Wilhelms-Universität Bonn

July 4, 2017

Abstract—Abstract

I. Introduction

II. Deep Learning - An Introduction

In the recent years, Deep learning has been making a big impact in various fields of computer science. This impact is profound in the perceptual learning task in the fields like computer vision, natural language processing, etc, [**?**]. Though the idea of training a multi layered neural network to perform function approximation is known to the research community for at least a couple of decades[**?**], due to the nature of training problem requiring large computational resources, the multi layered neural network remained less practical. With the advent of general purpose graphical processing units(GPGPUs) and the availability of training data, the neural networks are now a practical solution for many perceptual tasks. One of the early success of deep learning was in the field of image classification. In the annual ImageNet classification challenge [**?**], AlexNet [**?**] showed a remarkable improvement in the state-of-the-art accuracy. Within a few years, more sophisticated architectures like Google Inception Network [**?**], Deep Residual Network [**?**], etc., has improved the accuracy to be comparable with a human in that task. The generality of the neural network made it easily possible to be used for a wide variety of tasks. With more people working on deep learning and the ideas arising in solving on the problems being easily transferable, the deep learning is the state-of-the-art for many learning tasks. In the following section, we will briefly summarize the basics of deep learning.

The basic building block of a Multilayer network, or in more technical term Multilayer perceptron(MLP) is a perceptron. The preceptron **??** closely resembles a neuron in a human brain **??**. A preceptron takes a set of input values, computes a weighted sum of the inputs, and outputs a scalar value of a after applying an activation function (mostly non-linear). The weights for computing the weighted sum and the threshold in the activation function are initially set to random values and are learned from the training data. Mathematically, a perceptron with the weight matrix A and the threshold T for an step activation function, for the input vector x, outputs the following,

$$f(x) = \begin{cases} 1, & \text{if } Ax > c \\ 0, & \text{otherwise} \end{cases}$$

A layer has many number of neurons and many layers are stacked one on top of the other to form a MLP.
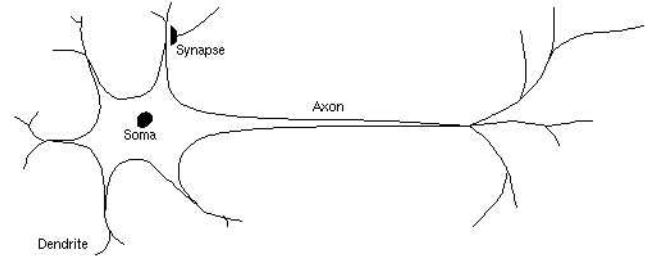

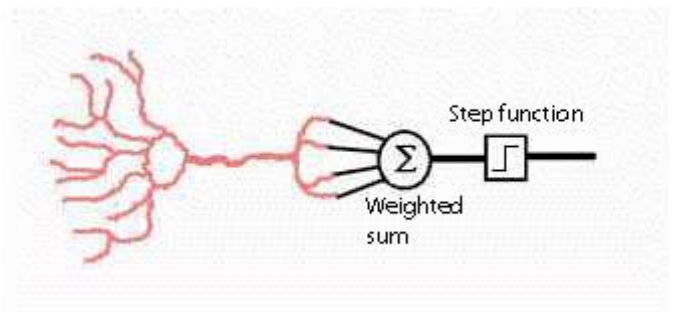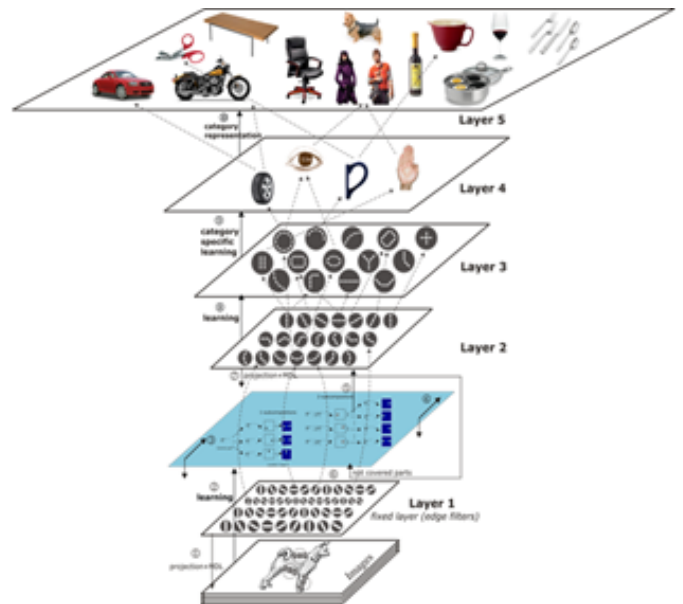Fig. 1: A biological neuron.

Fig. 2: An artificial neuron (perceptron)

Fig. 3: Layered representation in an MLP