**Convolutional Neural Networks**

**Introduction:**

The purpose of this document is to make you “Somberi’s” understand the how convolutional neural networks work and the math behind them if you are interested in learning about it. All the information provided in this document are referenced from research papers and from tutorials made by Geoffrey Hinton who is a well-known for his work on “Artificial Neutral Networks” and is also considered as the god father of deep learning.

1. **Preliminaries:**

Let’s start with some background knowledge that is necessary to understand how a convolutional neural networks work and the math behind them.

* 1. **Matrices Vector and Tensors:**

In simple words matrices represents a rectangle of numbers arranged in rows and columns. The row and column index is used to identify an element from the matrix.

Example: is a matrix

**Mathematical notations:**  we denote matrices as  H \* W. Here H represents the no of rows in the matrix and w represents the number of columns in the matrix.

Vector is a matrix with 1 column and many rows. Now comes the question why do we need vectors if we have arrays and matrices? We use vectors when we need the magnitude as well as the direction to get a complete idea about a variable. For example: Speed is a scalar quantity, because you don’t need the direction of motion to get an idea about what’s the current speed of a car. But, Velocity is a vector quantity because you need both the magnitude as well as the direction to calculate the velocity.

Example: is a vector.

**Mathematical notation:** we denote vectors as  H \* W. Here H represents the no of rows in the matrix and w is equal to 1.

Higher order matrices are called tensors. I couldn’t think of a way to explain tensors in simple terms so I am going to shift directly to the mathematical notation I the hope that you can understand when you see how a tensor is represented.

**Mathematical notation:** we denote tensors as  H \* W \* D. This is an order 3 tensor. It contains HWD elements. In simple terms consider a tensor as a thing that contains D channel matrices. Every channel is a matrix of size H \* W.

For example: Color image is a 3rd order tensor. If the color image is stored in RGB format then it has three channels with a matrix size of H \* W.

Now comes the question why do we use tensor when we can use vectors?

In early computer vision and pattern recognition, color images would be converted to gray scale (i.e. Black and white images) so that they can be stored in a matrix. The color information are lost during this transformation. But color is a very important feature in computer vision and pattern recognition. Due to these reasons tensors are used in CNN. The input, intermediate results, parameters are all tensors in CNN.

* 1. **Mathematical Concepts:**

**Note: This section is not important if you don’t want to understand the math behind CNN.**

The CNN learning process depends upon vector calculus and chain rule. Suppose Z is a Scalar (i.e. a single value or a number) and H is a vector. If z is a function of y meaning z = z(y) or in simple terms z varies according to the value of y. Then the partial derivatives of z with respect to y is given by.

Partial derivatives are used to denote how one variable varies with respect to the other. Chain rule is very simple

**Chain rule:**

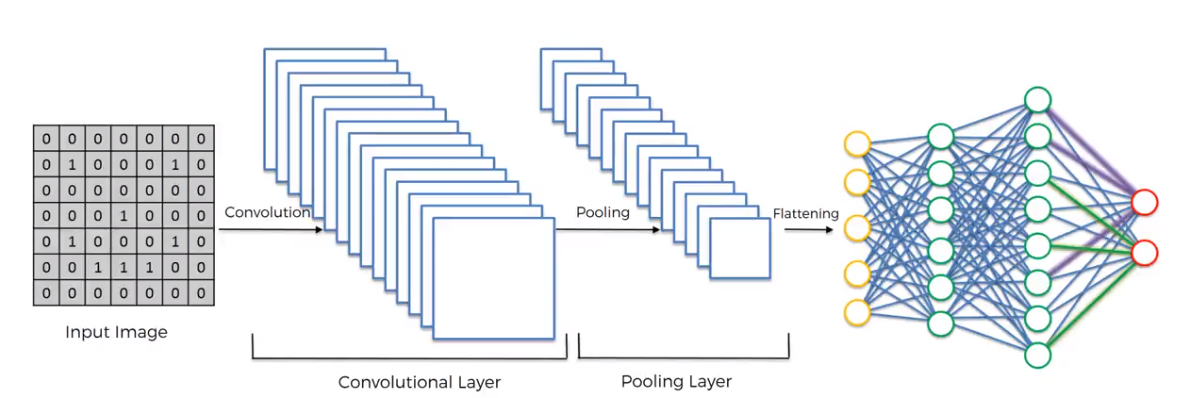
This equation is enough to understand chain rule. The definitions may confuse you that is why I have not provided any description about chain rule.

1. **Convolutional Neural Networks:**

In this section I am going to give you an abstract understanding about CNN and how it works. Detailed explanations will be given in the later sections.

* 1. **Architecture:**

The CNN normally takes a 3rd order tensor as input. The input is then passed to the convolution layer which performs the convolution function. The convolution function then produces an intermediate output which is then passed to the max pooling layer which reduces the size of the feature map.



Additional convolutional and max pooling layers can be added to the architecture. Then at last the intermediate results are flattened to get a feature vector. The feature vector is the passed to an ANN (Artificial Neural Network) which predicts the output.

* + 1. **What is CNN and Why do we need CNN:**

In deep learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep neural networks, most commonly applied to analyzing visual imagery. As business problems grew bigger the solution for those problems also grew more complex. Coding for a particular problem or an application became impossible as the problem was complex and also if a solution exists then that solution would be horrendous. For example: Object detection problems are very complex to code as there are many features that can be used to identify an object and those features can also be used to identify other objects as well. The CNN way of approaching this problem is to allow the model to find the features that pertains to that object automatically by comparing it with the actual or the desired output. The model learns by itself using mathematical operations which enable the model to adapt even if the features are changed later.

You may think that CNN are more complex than actually coding for a problem. But CNN’s are actually easy to understand and there is no need to know about the mathematics behind it.