Random notes about Math for Deep Learning Models

Konstantin Burlachenko (burlachenkok@gmail.com)

Last update: 18 APR 2020

Contents

1 Introduction

This document is suggested to be evolved in time and contains some general math things in which DL aspects is leveraging on or heuristically or non-heuristically.
It's only in the beginning phase and will be evolved in time.

2 Differentiation

2.1 Introduction about Differentiation

This day's people talk about chain rule when prediction schema $F(x;\theta) : X \to Y$ constructed as composition of several functions.

2.2 Differentiation Definition

Function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable if for arbitrarily $\Delta x \in \mathbb{R}^n$ which elementwise consist of some components, without any inter-depency we can state thate

$$f(x + \Delta x) - f(x) = A\Delta x + \alpha(\Delta x)|\Delta x|$$

Where:

1. $\alpha(\Delta x)$ is infinite small function i.e. $lim_{\Delta x \to 0}\alpha(\Delta x) = 0$
2. A is a rectangular matrix and its items should not depend on specific direction or values of Δx
3. $\alpha(\Delta x)$ is infinite dimensional function relative to Δx

Differentiation is a local property of the function. If function is differentiable in all points in some set χ then people say that function is differentiable in the whole set χ.

2.3 Some Theorems in context of Differentiation

Theorem 2.1 (Partial Derivatives)**.** *If function f is differentiable in point then all partial derivatives of the function exist and equal to corresponded elements of the matrix A.*

To derive this theorem you should vary $\Delta x = [0, 0, ..., 0, dx_i, 0, ..., 0]^T$. This simple variation in input will allow eliminating every from A, except i-th column.

Theorem 2.2 (Continuity Property)**.** *If function f is differentiable in point then function is continuous in the point.*

Theorem 2.3 (Differentiability from existing partial derivatives)**.** *If function f has all partial derivatives in point a and they are continuous in functions in point a then function is differentiable in point a.*

Theorem 2.4 (Differentiability of function composition)**.** *Also known as the chain rule.*

1. *If function f is differentiable in point of it's domain a.*
2. *If function g is differentiable in point of it's domain b.*
3. *If $b = f(a)$*

Then function $z = g(f(x))$ is differentiable in points a and $z' = g' \cdot f'$

3 What is convolution in mathematics

3.1 Convolution definition

I think you know that convolution is well-defined operation:

$$f * g(x) = \int_{-\infty}^{+\infty} f(y)g(x-y)dy \tag{1}$$

Sometimes limits of integration in (1) can be reduced to some subinterval of $[-\infty, +\infty]$.

3.2 Convolution commutativity

$$\begin{multline} f * g(x) = \int_{-\infty}^{+\infty} f(y)g(x-y)dy = \\ |x-y=z, dy=-dz| = \\ \int_{+\infty}^{-\infty} f(x-z)g(z)(-dz) = \\ \int_{-\infty}^{+\infty} g(z)f(x-z)dz = g * f(x) \end{multline} \tag{2}$$

So we have proved that $f * g = g * f$

3.3 Time invariant property of convolution

$$\tau(f*g,b) = f*g(x-b) = \int_{-\infty}^{+\infty} f(y)g((x-b)-y)dy =$$
$$|g(x-b) = \tau(g,b)$$
$$\int_{-\infty}^{+\infty} f(y)\tau(g,b)(x-y)dy =$$
$$f*\tau(g,b) \quad (3)$$

So we have proved that $f*g = g*f$

3.4 Some Convolution Algebraic Properties

$$f*g = g*f \quad (4)$$
$$(f*g)*h) = f*(g*h) \quad (5)$$
$$f*(g+h) = f*g+f*h \quad (6)$$
$$\frac{d^k(f*g)}{dx^k} = f*\frac{d^k g}{dx^k} \quad (7)$$
$$f*\delta(a) = f(a) \quad (8)$$

4 What is cross-correlation in math

4.1 Cross-Correlation definition

If consider functions $\mathbb{R} \to \mathbb{R}$ then value of cross-correlation for them is **defined point-wise for specific value x** as the following

$$f \star g(x) = \int_{-\infty}^{+\infty} f(y)g(x+y)dy \quad (9)$$

If compared with convolution there is no "flipping and dragging" of convolution kernel in algebraic definition.

What Neural Network community mean by Convolution is some form of Cross-Correlations

4.2 Cross-Correlation connection with convolution

Let's define function $f^{-} := f(-x)$. This operation takes a function and reverse it w.r.t. to function value axis.

$$
\begin{aligned}
f \star g(x) = \int_{-\infty}^{+\infty} f(y)g(x+y)dy = \\
|y = -z, dy = -dz| = \\
\int_{+\infty}^{-\infty} f(-z)g(x-z)(-dz) = \\
\int_{-\infty}^{+\infty} f(-z)g(x-z)dz = \\
\int_{-\infty}^{+\infty} f^{-}g(x-z)dz = f^{-} * g
\end{aligned}
\tag{10}
$$

So we derived $f \star g = f^{-} * g$

4.3 Some Cross-Correlation Properties not necessary to hold

Convolution behaves like multiplication, but as you see $f \star g = f^{-} * g$ So due to that different algebraic properties is not necessary to hold. List of some properties which in general does not hold

$$f \star g = f^{-} * g = g * f^{-} \neq g^{-} * f = g \star f \tag{11}$$

$$(f \star g) \star h \neq f \star (g \star h) \tag{12}$$

5 To be continued

Deep Learning models evolve in time and attack composition function representation from different ways and learning too.