Neural Networks

Kieran Harvie

This document is still under construction.
Changes in content, structure, and readability are expected.

Contents

Introduction

Derivatives

General

2.1 A Tale of Two Derivative

Partial and total. Chain rule. ∇ operator maybe an appendix for vectors?

$$\frac{d\,f}{d\,t} = \sum_i \frac{\partial f}{\partial x_i}\frac{d\,x_i}{d\,t} = \nabla f \cdot \left(\frac{d\,x_0}{d\,t}, \frac{d\,x_1}{d\,t}, \frac{d\,x_2}{d\,t}, \dots\right)$$

2.2 Dual Numbers

Like how we get Complex numbers from Real numbers by adjoining a new number i such that $i^2 = -1$ we get the Dual numbers by adjoining a new number ϵ, called the infinitesimal, such that $\epsilon^2 = 0$.

Some properties aren't that special:

$$(x + x'\epsilon) + (y + y'\epsilon) = (x + y) + (x' + y')\epsilon$$

But consider what happens for multiplication and division:

$$\begin{aligned}
(x + x'\epsilon)(y + y'\epsilon) =& xy + (x'y + xy')\epsilon + x'y'\epsilon^2 \\
=& xy + (x'y + xy')\epsilon \\
\frac{x + x'\epsilon}{y + y'\epsilon} =& \frac{(x + x'\epsilon)(y - y'\epsilon)}{(y + y'\epsilon)(y - y'\epsilon)} \\
=& \frac{xy + (x'y - xy')\epsilon - x'y'\epsilon^2}{y^2 - y'^2\epsilon^2} \\
=& \frac{x}{y} + \frac{x'y - xy'}{y^2}\epsilon
\end{aligned}$$

Define: $f\left(\langle u, u'\rangle, \langle v, v'\rangle\right) = \langle f(u,v), \nabla f(u,v) \cdot (u', v')\rangle$
Because:

$$f\left(\left\langle u, \frac{du}{dt}\right\rangle, \left\langle v, \frac{dv}{dt}\right\rangle\right) = \left\langle f(u,v), \nabla f(u,v) \cdot \left(\frac{du}{dt}, \frac{dv}{dt}\right)\right\rangle$$
$$= \left\langle f(u,v), \frac{d}{d\,t} f(u,v)\right\rangle$$

(Because of the similar form this argument also works with partials).

2.2.1 Historical Note

This is actually close to how historic calculus Transcendental law of homogeneity only keep the lowest term. Adequality.

2.3 Automatic Differentiation

2.3.1 Forward Accumulation

$$f(\langle x, 1\rangle, \langle y, 0\rangle) = \left\langle f(x,y), \frac{\partial}{\partial x} f(x,y)\right\rangle$$

2.3.2 Reverse Accumulation

$$f\left(\left\langle x, \frac{dx}{df}\right\rangle, \left\langle y, \frac{dy}{df}\right\rangle\right) = \left\langle f(x,y), \frac{df}{df}\right\rangle = \langle f(x,y), 1\rangle$$

Adjoint:

$$\bar{x} = \frac{\partial f}{\partial x}$$

Optimization

3.1 Gradient Descent

3.2 Newton's Method

Like a dynamic step size since we have this info? Only when root finding. (Use duel numbers).

Using error l_2 all the functions are twice differentiable so you can do the Lagrange form of the remainder thing. Also only one root so how does that work with basins

$$x_{n+1} = x_n - \frac{F(x_n)}{|\nabla F(x_n)|^2} \nabla F(x_n)$$

Nodes

Transfer (SUM) Activator (Function (logistic)) Loss/Cost (Error)

4.1 Layered Structure

Back propagate but only keep one row in memory at a time.

Appendix

A.1 Logistic Function