INTRODUCTION TO CALCULUS: MULTIVARIATE CALCULUS, DIFFERENTIATION AND INTEGRATION RULES

1. Partial Derivatives

So far, we have encountered the derivative of one variable. However, when dealing with Machine Learning or other Data Analytics problems, we are frequently dealing with multiple variables. For example, even in the *auto* data set that we looked at in our assignments, the response variable *mpg* was described as a function of 4 other variables.

Formally, let us assume that we have a function that is defined as: $f : R^n \to R$. That is, it is a function whose Domain is an n-dimensional real-valued vector and its range is a scalar, also real-valued. For example, the *auto* data takes a 4-dimensional vector as an input and produces a 1-dimensional *mpg* estimate as the output value. Then, we can define derivatives of the function in terms of each of the variables. We can define $f'(x_i)$ as the derivative of the function f in terms of the ith dimension, represented by the variable x_i. This is defined as the (usual) derivative of the function of a single variable assuming that all other variables are constant. For example if we have a function of two variables: $f(x_1, x_2)$ as below:

$$f(x_1, x_2) = a_1x_1 + a_2x_2 + a_{12}x_1x_2 \tag{1}$$

where the a's are constants. Its partial derivative along x_1 is

$$\frac{\partial f}{\partial x_1} = a_1 + a_{12}x_2 \tag{2}$$

Similarly, the partial derivative along x_2 is

$$\frac{\partial f}{\partial x_2} = a_2 + a_{12}x_1 \tag{3}$$

The partial derivative is indicated by a special symbol ∂ to denote that the resulting function represents only a part of the total derivative and it is in the direction of the dimension (x_1 or x_2) along which it is taken. That is, it is indicating the nature of the slope of the function along the direction represented by x_1 or x_2, respectively. Partial derivative are derivatives along a specific direction. You can take derivatives along any direction and the partial derivatives just happen to be special directional derivatives where those directions line up with the axis of the various dimensions.

2. Gradient

Given a function $f : R^n \to R$ the gradient is just another function from $R^n \to R^n$ where each dimension is simply the partial derivative along that direction. It is represented by

Common Function	Form	Differential Form
Constant	$f(x) = c$	$f'(x) = 0$
Identity	$f(x) = x$	$f'(x) = 1$
Square	$f(x) = x^2$	$f'(x) = 2x$
Square root	$f(x) = \sqrt(x)$	$f'(x) = \frac{1}{2}x^{-\frac{1}{2}}$
General form of the above	$f(x) = x^n$	$f'(x) = nx^{n-1}$
Exponential	$f(x) = e^x$	$f'(x) = e^x$
	$f(x) = a^x$	$a^x(ln(a))$
Logarithm	$f(x) = ln(x)$	$f'(x) = \frac{1}{x}$
Trigonometry	$f(x) = \sin(x)$	$f'(x) = \cos(x)$
	$f(x) = \cos(x)$	$f'(x) = -\sin(x)$

Rule	Function	Differential Form
Multiplication by constant	$cf(x)$	$cf'(x)$
Sum of Functions	$f(x) + g(x)$	$f'(x) + g'(x)$
Product of Functions	$f(x)g(x)$	$f'(x)g(x) + g'(x)f(x)$
Reciprocal	$1/f(x)$	$-f'(x)/f^2(x)$
Quotient	$f(x)/g(x)$	$(f'(x)g(x) - g'(x)f(x))/g^2(x)$
Chain rule	$f(g(x))$	$f'(g(x))g'(x)$

the symbol Δ. In other words,

$$\Delta f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{4}$$

For a one dimensional function $f : R \rightarrow R$, the gradient, the derivative and the slope are all synonymous. The gradient in multiple dimensions is simply the slope of the function in those dimensions.

3. Differentiation Rules

The tables contain some differentiation rules that help you work out derivatives of functions. In addition to some common rules above, which can be proved using limits as we did in the previous module, here are some rules that allow you to work with complex functions. Let's take a look at the chain rule in the second table. When we can express a function as a composition of several functions, we can apply the chain rule to take the derivative easily. If you examine the chain rule, it makes intuitive sense. For example, consider the function $h(x) = a_2 \sin(x)^2 + a_1 \sin(x) + a_0$. You can consider this as a function of a new variable $y = \sin(x)$. This is just a polynomial function of y and y happens to be a function of x. So, when we differentiate such functions, we first take the derivative with respect to the new variable y and then take the derivative of y with respect to x.

Form	Integral Form
$\int a dx$	$ax + C$
$\int 2x dx$	$x^2 + C$
$\int x dx$	$x^2/2 + C$
$\int x^n dx$	$x^{n+1}/(n+1) + C$
$\int e^x dx$	$e^x + C$
$\int (1/x) dx$	$ln\|x\| + C$

Rule	Function	Integral Form
Multiply by Constant	$\int cf(x)dx$	$c\int f(x)dx$
Sum of Functions	$\int [f(x) + g(x)]dx$	$\int f(x)dx + \int g(x)dx$

4. Integration Rules

Given the differentiation rules, we can derive the integration rules from them. Some common ones are shown in the table. We also can derive some common rules to deal with complex functions and their integrals. Let's now see a special rule called *integration by parts*. This comes from the product rule of differentiation: $(fg)' = f'g + g'f$. Now, if you integrate both sides of this equation, we get

$$\begin{aligned} \int (fg)' dx &= \int fg' dx + \int f' g dx \\ fg &= \int fg' dx + \int f' g dx \\ \int fg' dx &= fg - \int f' g dx \end{aligned} \tag{5}$$

The last form $\int fg'dx = fg - \int f'gdx$ is frequently referred to as integration by parts. Given a product of two functions, you consider one of the function to be f and the other to be g', or the derivative of g. Choosing f and g appropriately can cause these integrals to simplify and factor out easily.