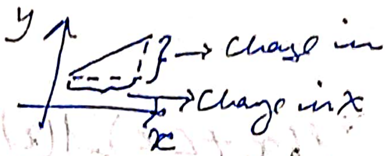
Calculus

ML Use Cases

* Numerical Optimization
* Gradient Computations
* Probability Density Functions
* Variational Inference and Related Techniques (Later-3 yr)

Derivatives

Slope = $\frac{\text{Change in } x}{\text{Change in } y}$



The derivative of a function $y = f(x)$ with respect to x is defined as

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

where x value is increased

the old value of x.

~~provided prod~~

provided that the limit exists

Derivative of $y = f(x)$ with respect to x is represented as dy/dx or $f'(x)$.

* $y = f(x) = 3x$

Initially $x = 5 \Rightarrow y = 15$

Increase x by 0.001 what will be the corresponding change in y?

~~det~~ $\frac{\Delta y}{\Delta x} = ?$

for y, if the initial value of x is 5 and $y = 3x$,

$y = 3 \times 5 = 15$

So initially, $y = 15$

Now, if we increase x by 0.001, we get

$x = 5 + 0.001 = 5.001$

Using the function $y = 3x$, we can find the corresponding value of y:

$$y = 3 \times 5.001 = 15.003$$

$\therefore$ The final value of y is 15.003 when x is increased by 0.001

Now, $\frac{\Delta y}{\Delta x} = ?$

$$\frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{(3(x+\Delta x) - 3x)}{\Delta x}$$

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{3\Delta x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} 3$$

As Δx approaches 0, the expression approach 3.

So, $\frac{dy}{dx} = 3$. This means that for every unit increase in x, y increases by 3.

$g(x) = \sin x$

$h(x) = 3x$

$y = f(x) = g(h(x))$

$= g(3x)$

$= \sin(3x)$

$y = a(x) = \cancel{h(g(x))}\ h(g(x))$

$= h(\sin x)$

$= 3 \sin x$

Chain Rule

$y = f(x) = g(h(x))$

$$\frac{dy}{dx} = f'(x) = \frac{dg(h(x))}{d(h(x))} \times \frac{d(h(x))}{dx}$$

$y = f(x) = \sin 3x$

$$\frac{dy}{dx} = f'(x) = \frac{d(\sin 3x)}{d(3x)} \times \frac{d(3x)}{dx}$$

$= \cos(3x) \times 3$

$= 3\cos 3x$

$\frac{d(3x)}{dx} = 3.$

$\sin x = \cos x$

$3x = 3$

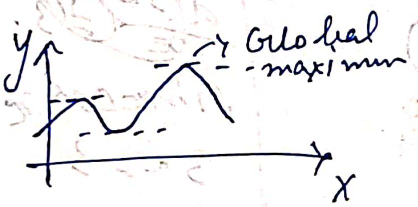
Maxima and Minima using Derivatives

In a smoothly changing function a maximum or minimum is always where the function flattens out (except for a saddle point)

* Saddle Point :- where the derivative will be zero but the point is neither ~~or~~ Local Maximum and or Local minimum

Smoothly changing continuous function flattens when it becomes parallel to x axis.

* Global maximum → highest point of local maximum



* How do we know if its maximum or minimum?
if $f''(x_0) < 0 \Rightarrow x_0$ is point of maxima
if $f''(x_0) > 0 \Rightarrow x_0$ is point of minima.

* Finding Minima & Maxima

$y = (x-1)(x-2)(x-3)$

$(x-1)$ and $(x-2)$

$\Rightarrow x^2 - 2x - x + 2 = x^2 - 3x + 2$

$(x^2 - 3x + 2)(x-3)$

$\Rightarrow x^3 - 3x^2 - 3x^2 + 9x + 2x - 6$

$\Rightarrow x^3 - 6x^2 + 11x - 6$

$f'(x) = 0 \Rightarrow 3x^2 - 12x + 11 = 0$

derivative of x^3 is $3x^{3-1} = 3x^2$

$6x^2$ is $2.6x^{2-1} = 12x$

$11x$ is $11 = 11$

-6 is $0 = 0$.

$3x^2 = 6x$
$12x = 12$
$11 = 0$.
$\cancel{5+1.7}$
2

$f'(x) = 0 \Rightarrow 3x^2 - 12x + 11 = 0$.

$\Rightarrow 6x - 12$

quadratic formula: $ax^2 + bx + c = 0$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$a = 3$, $b = 12$ & $c = 11$.

$\cancel{x = +12}$

$$x = \frac{12 \pm \sqrt{(-12)^2 - 4(3)(11)}}{2(3)}$$

$$x = \frac{12 \pm \sqrt{144 - 132}}{6}$$

$$x = \frac{12 \pm \sqrt{12}}{6}$$

$$x_1 = \frac{12 + \sqrt{12}}{6} \qquad x_2 = \frac{12 \pm \sqrt{12}}{6}$$

$x_1 = \frac{12+\sqrt{12}}{6}$ $x_2 = \frac{12-\sqrt{12}}{6}$

$\sqrt{12} = 3.46$

$x_1 = (12+3.464)/6$

$= 2.577$

~~$x_2 = 1.423$~~

$x_2 = (12-3.464)/6 = 1.423$

1.423
$\times 6$
8.538

12.000
-8.538
3.462

$f''(x) = 6x - 12$

~~$f''(1.42) = 6x$~~

$f''(2.57) = 6 \times 2.57 - 12$

$f''(2.57) = 15.42 - 12$

$\therefore$ the value of the expression $6x - 12$ when x is 2.57 is:

$f(2.57) = 3.42$ (+ve). So $x_0 > 0$, it's a local minima.

$f''(x) = 6x - 12$

$f''(1.423) = 6 \times 1.423 - 12$

$f''(1.423) = -3.462$ (−ve). So $x < 0$, it's a local maxima

Partial Derivatives

In functions with two or more variables, the partial derivative is the derivative of one variable with respect to the others.

$z = f(x, y)$

∴ If we change x, z will change.
if we change y, z ″ ″
if we change both z will change.

$z = f(x, y)$

Now if we change x, but hold all other variable constant, how does $f(x, y)$ change?
That's one partial derivative.
The next variable is y.

~~if we change~~

$f(x, y) = 2x^2 + 3xy + 6x + 7y$

with respect to x, we will not change y.
y will be constant.

$\frac{\partial f}{\partial x} = 4x + 3y + 6$

We're taking $3y$ as constant. x derivative is 1
$7y \Rightarrow$ constant
$6x \Rightarrow x = 1$

$\frac{\partial f}{\partial x} = 4x + 3y + 6.$

$\frac{\partial^2 f}{\partial x^2} = 4$ #$3y$ & 6 is constant.

$y = x$

$\frac{dy}{dx} = 1.$

$2x^2$, $a = 2$
$n = 2$

$\frac{d}{dx}(2x^2)$

$= 2 \cdot 2x^{2-1}$

$= 4x$

Partial Derivatives

In functions with two or more variables, the partial derivative is the derivative of one variable with respect to the others.

$z = f(x, y)$

$\therefore$ If we change x, z will change.
if we change y, z " "
if we change both z will change.

$z = f(x, y)$

Now if we change x, but hold all other variable constant, how does $f(x, y)$ change?
That's one partial derivative.
The next variable is y.

~~if we change~~

$f(x, y) = 2x^2 + 3xy + 6x + 7y$

with respect to x, we will not change y.
y will be constant.

$\frac{\partial f}{\partial x} = 4x + 3y + 6$

we're taking $3y$ as constant. x derivative is 1
$7y \Rightarrow$ constant
$6x \Rightarrow x = 1$

$\frac{\partial f}{\partial x} = 4x + 3y + 6.$

$\frac{\partial^2 f}{\partial x^2} = 4$ # $3y$ & 6 is constant.

$y = x$

$\frac{dy}{dx} = 1.$

$2x^2$, $a = 2$
$n = 2$

$\frac{d}{dx}(2x^2)$

$= 2 \cdot 2x^{2-1}$

$= 4x$

Derivative with respect to y

$2x^2$ & $6x^{\text{x}}$ is constant

$y = 1$.

$$\Rightarrow 2x^2 + 3xy + 6x + 7y$$

$$\cancel{\frac{\partial f}{\partial y}} \frac{\partial f}{\partial y} = 3x + 7$$

$3x$ is constant

$7 \Rightarrow 0$

$$\frac{\partial^2 f}{\partial y^2} = 0$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = 3.$$

Jacobian matrix

Consider the function $f: R^2 \to R^2$ given by

$$f(x,y) = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix}$$

Then we have

$$f_1(x,y) = x^2 y$$

and

$$f_2(x,y) = 5x + \sin y$$

~~and the Jacobian matrix of F is~~

& the Jacobian matrix of F is

$$Jf(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos y \end{bmatrix}$$

& the Jacobian determinant is

$$\det(Jf(x,y)) = 2xy\cos y - 5x^2.$$

$\frac{d(x^2)}{dx} y + x^2 \times \frac{d(y)}{dy}$

$= 2x \times y + x^2 \times 1$

$(x^2 y)$

$= \frac{d}{dx}(x^2) \cdot y$

$y = 2xy$

$\cancel{x^2 y = 5x}$

$\cancel{\frac{d(x^2)}{dy} = 0}$

$x^2 y$

$x^2 \cdot \frac{dy}{dy} + y \cdot \frac{dx}{dy}$

$\frac{dy}{dy} = 1$

$\frac{d(x^2)}{dy} = 2x$

Definition: It's a way to describe how changes in one set of variables affect changes in another set of variables.

vector-valued function $f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix}$

where $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is the vector of input variables, then the Jacobian matrix J is defined as:

$$\cancel{J = \begin{matrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \partial f_2 & & \end{matrix}}$$

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

The Jacobian matrix provides information about how small changes in the input variable x lead to changes in the output variables $f(x)$. ~~which is useful in various m~~

* Important in optimization (gradient descent), backpropogation (used in training neural networks), and sensitivity analysis.

Derivatives Formulas

Constant Rule: $\frac{d}{dx}(c) = 0.$

" Multiple Rule: $\frac{d}{dx}[cf(x)] = cf'(x)$

Power Rule: $\frac{d}{dx}(x^n) = nx^{n-1}$

Sum Rule: $\frac{d}{dx}[f(x) + g(x) = f'(x) + g'(x)$

Difference Rule: $\frac{d}{dx}[f(x) - g(x) = f'(x) - g'(x)$

Product Rule: $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$

Quotient Rule: $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

Chain Rule: $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$

$\frac{d}{dx}(\sin x) = \cos x$ | $\frac{d}{dx}(\cos x) = -\sin x$

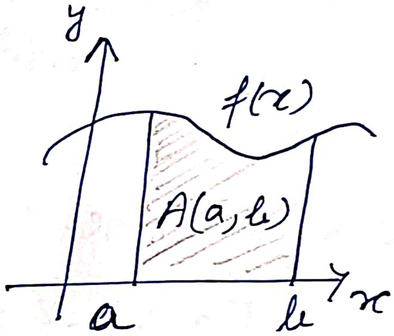
$\frac{d}{dx}(\tan x) = \sec^2 x$ | $\frac{d}{dx}(\cot x) = -\csc^2 x$

$\frac{d}{dx}(\sec x) = \sec x \tan x$ | $\frac{d}{dx}(\csc x) = -\csc x \cot x$

Definite Integrals

The integral of $f(x)$ corresponds to the computation of the area under the graph of $f(x)$. The area under $f(x)$ between the points $x=a$ & $x=b$ is denoted as follows:

$$A(a,b) = \int_a^b f(x)\,dx$$



Sigmoid function: $\sigma(x) = \frac{1}{1+e^{-x}}$

where e is Euler's no. (approx. 2.71)

sigmoid fn. takes any i/p x and maps it to a value between 0 and 1. This property make it useful for models we need to predict probabilities. Ex- in binary classification problems, the o/p of a neural network can be passed throug a sigmoid fn. to get probability score between 0 and 1 indicating the likehood of a certain class.

~~However,~~

However, the sigmoid fn. has some drawbacks, particularly in deep neural networks.

* Saddle point :- Point where fuction neither have maximum value nor minimum value.