Calculus For Machine learning & DATA SCIENCE

What is a derivative?

derivatives can find the max & min value of a function.

Calculus

Derivative

How fast a quantity is changing at a specific point

If a function $F(x)$ gives you a value based on input x, then the derivative $F'(x)$ tells you:

- The rate of change of $F(x)$
- Or how much $F(x)$ increases or decreases as x changes a little bit.

Why Important for ML?

To optimise functions, to maximise or minimise it.
When you find a model that fits your dataset the best, you calculate a Loss function & minimise it.

2 houses

1 house
1 Bedroom
150,000

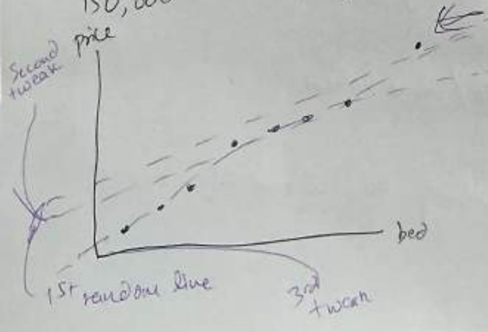
2 house
2 bedroom
250000

> Find price of house by the no. of bedrooms

No. of bedrooms	Price of house
1	150
2	250
3	300
4	600
5	650
6	750
7	800
8	
9	
10	1,05000

??

use ML model to predict



The idea here is that the ML model takes all data & begins a process called model training → like the engine (smart engine) simply a line

It tweaks itself until it can give a good prediction.
This problem is called a linear regression.

Let's say you go 100km/h.

so avg velocity → 100km/h.

but you could have been going at 200km/h & stopped in between

averages out

so the question is at a given instant what's the velocity at the given instant → instantaneous velocity → this is derivative

Instantaneous rate of change of function.

Speed = d/t

let's say speedometer broke. [look in slide on table].

The car is not moving at a constant speed, if you measure distance travelled every 5 seconds

Can you use this to find velocity at t = 12.5 seconds.?

Velocity = distance travelled / time taken.

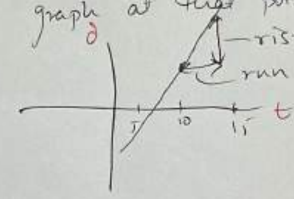
No. But you can find avg velocity in interval 10-15 seconds.

80/5 = 16

[illegible]

If we draw a graph of the points on table [in slide]
zoom into the part 10 - 15 seconds.

The slope of the graph at that point is the ~~avg~~ velocity at that point.



$$\text{Slope} = \frac{\text{rise}}{\text{run}}$$

In this case

rise is change in distance
run is change in time
> $\text{slope} = \frac{\text{rise}}{\text{run}} \Rightarrow \frac{\text{change in distance}}{\text{change in time}}$

Synonomous to velocity

$$\text{slope} = \frac{d(15) - d(10)}{t(15) - t(10)} \Rightarrow \frac{202 - 122}{15 - 10} \Rightarrow \text{slope} = \frac{80m}{5s} \Rightarrow 16m/s.$$

This is a good prediction for v at 12.5s.

Now can we find a better prediction? → If we had more data around 12.5s yes. Now let's say we take measurements every 1s.

[Data is in table in slide]

$d(12) \Rightarrow 155 \quad d(13) = 170$

$\frac{170 - 155}{13 - 12} \Rightarrow 15 m/s$ → more accurate guess.

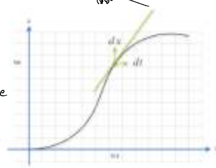
$$\text{Slope} \Rightarrow \frac{\Delta d}{\Delta t} \Rightarrow \frac{d(13) - d(12)}{13 - 12} \Rightarrow \frac{170 - 155}{13 - 12} \Rightarrow \text{slope} = 15 m/s$$

DeepLearning.AI

The derivative

$$\frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \longrightarrow \frac{dx}{dt}$$

This slope is the instantaneous rate of change.

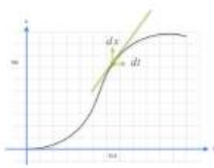


So you can see here how the small adjustments are made by the model to reach the potential derivative

DeepLearning.AI

The derivative

$$\frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \quad \frac{\Delta x}{\Delta t} \longrightarrow \frac{dx}{dt}$$



Derivative

The slope of the tangent at that particular point

DeepLearning.AI

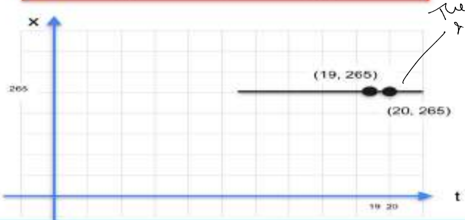
DeepLearning.AI

Derivatives and Optimization

Slopes, maxima, and minima

Zero Slope

t (s)	x (m)
19	265
20	265



The line is horizontal & a horizontal line has a slope of 0.

$$slope = \frac{\Delta x}{\Delta t}$$

$$slope = \frac{x(20) - x(19)}{20 - 19}$$

$$slope = \frac{265m - 265m}{1s}$$

$$slope = \frac{0}{1}$$ ← No rise in distance

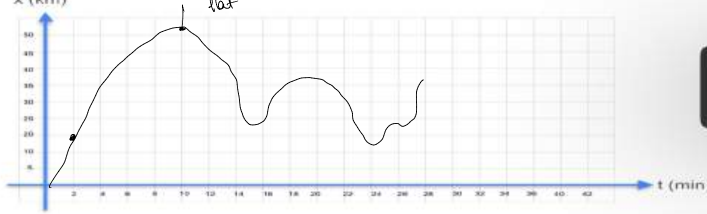
$$slope = 0\ m/s$$

Minima and Maxima

(I'm going to draw a quiz velocity of left car on this graph.

Quiz! Where was the velocity of the car zero?

- ☑ t=10
- ☑ t=16
- ☐ t=24
- ☑ t=22
- ☑ t=34
- ☑ t=40



looks pretty flat

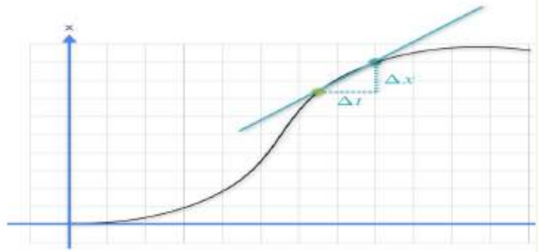
Wherever there were horizontal to tangent

Derivatives

$$\text{slope} = \frac{\text{change in distance}}{\text{change in time}}$$

$$\text{slope} = \frac{\Delta x}{\Delta t}$$

$$\text{slope at a point} = \frac{dx}{dt}$$

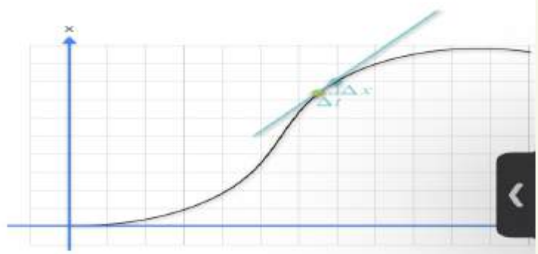


Derivatives

$$\text{slope} = \frac{\text{change in distance}}{\text{change in time}}$$

$$\text{slope} = \frac{\Delta x}{\Delta t}$$

$$\text{slope at a point} = \frac{dx}{dt}$$



DeepLearning.AI

Derivatives

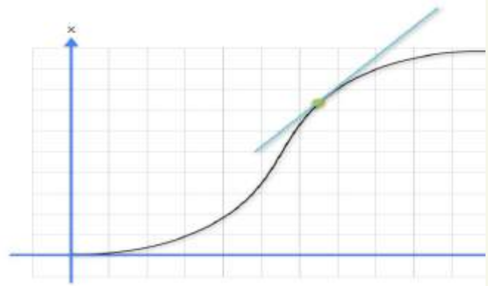
slope = change in distance

Derivatives

$$\text{slope} = \frac{\text{change in distance}}{\text{change in time}}$$

$$\text{slope} = \frac{\Delta x}{\Delta t}$$

$$\text{slope at a point} = \frac{dx}{dt}$$

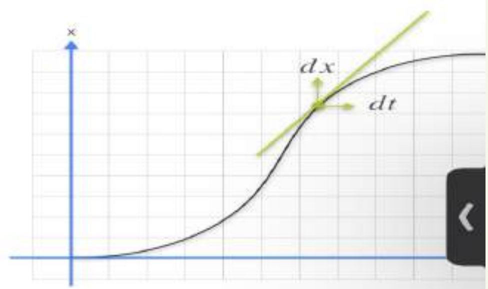


Derivatives

$$\text{slope} = \frac{\text{change in distance}}{\text{change in time}}$$

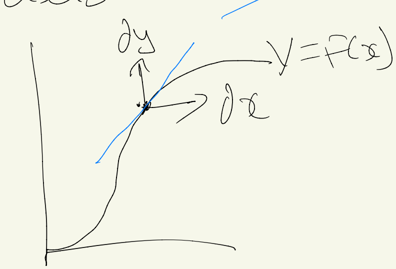
$$\text{slope} = \frac{\Delta x}{\Delta t}$$

$$\text{slope at a point} = \frac{dx}{dt}$$



DeepLearning.AI

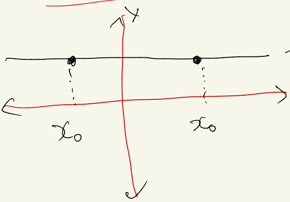
Let's now make horizontal axis as x & vertical axis as y. slope $= f'(x)$ $\quad y = f(x)$



derivative of f can be represented as $f'(x)$ [Lagrange's notation]

$\frac{dy}{dx} = \frac{d}{dx} f(x)$ → Leibniz's notation

Derivative of a Constant

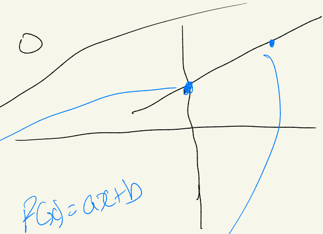


$y = f(x) = c$

$$\text{slope} = \frac{dy}{dx} = \frac{\Delta y}{\Delta x} = \frac{c-c}{x_1 - x_0} \Rightarrow 0$$

so derivative $= 0$.

Derivative of a line



$(x, ax+b)$

$f(x) = ax+b$

$(x+\Delta x, a(x+\Delta x)+b)$

so $f(x) = ax+b$ — $f'(x) = a$

$$\frac{\Delta y}{\Delta x} = \frac{\text{rise}}{\text{run}} = a$$

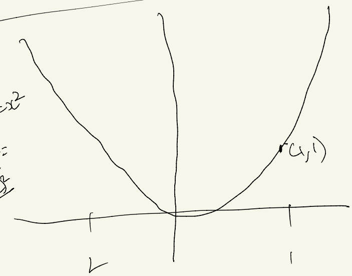
now.

$$\frac{df}{dx} = \frac{\Delta y}{\Delta x} = \frac{a(x+\Delta x) + \cancel{b} - (ax + \cancel{b})}{\Delta x}$$

$$\Rightarrow a \cdot \Delta x / \Delta x \Rightarrow \underline{a}$$

$y = f(x) = x^2$

$\text{slope} = \frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^2 - (x)^2}{\Delta x}$



Now if you begin with a big shift → say (1,1) to (2,4) →

$\text{slope} = \frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^2 - (x)^2}{\Delta x}$ → $\Delta x = 2-1 = 1$

$(x+\Delta x)^2 \Rightarrow (1+1)^2 - ?$ → $4-1$ $= 3$

If you see this qualities $f(x)=x^2$, $f(1)=1$, $f(2)=4$

$\Delta x = 1$
$\Delta f = 3$
$\text{slope} = 3/1 = 3$

Now let's make interval smaller. (1,1), (1.5, 2.25)

$$\Rightarrow \frac{(1+0.5)^2 - 1^2}{0.5} \Rightarrow \frac{1.25}{0.5} = 2.5$$

Δf, Δx

Now, $1/1000$

slope → $2.001 \approx 2$.

the derivative of $x^2 = 2x$ so for $x=1$ slope $= 2$

Let's derive

$$\frac{\Delta f}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x} \Rightarrow \frac{df}{dx} = \frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^2 - x^2}{\Delta x} = \frac{\cancel{x^2} + 2x\Delta x + (\Delta x)^2 - \cancel{x^2}}{\Delta x}$$

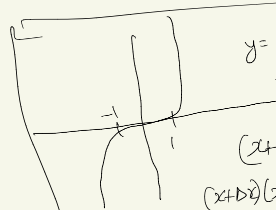
$$\Downarrow$$

$$2x \Leftarrow 2x + \Delta x \Leftarrow \frac{(2x+\Delta x)\cancel{\Delta x}}{\cancel{\Delta x}}$$

Now Δx when it goes to 0,
so $\Delta x \approx 0$

$y = \text{cubic} = f(x) = x^3$

$\text{slope} = \frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^3 - (x)^3}{\Delta x}$



$(x+\Delta x)\,(x+\Delta x)^2 - x^3$

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

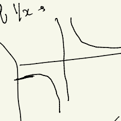
$\Rightarrow x^3 + 2x^2\Delta x + \Delta x^2 x + x^2\Delta x + 2x\Delta x^2 + \Delta x^3 - x^3$

$\Rightarrow \cancel{x^3} + \Delta x^3 + 3x(\Delta x)^2 + 3x^2(\Delta x) - \cancel{x^3}$

$\Rightarrow \frac{\Delta x(\Delta x^2 + 3x\Delta x + 3x^2)}{\cancel{\Delta x}} \Rightarrow \Delta x^2 + 3x\Delta x + 3x^2$

Now Δx reaches 0. $\Rightarrow 3x^2$

Derivative of $1/x$ →



$y = f(x) = x^{-1} = 1/x$

$\text{slope} = \frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^{-1} - (x)^{-1}}{\Delta x}$

$\Rightarrow \frac{\frac{1}{x+\Delta x} - \frac{1}{x}}{\Delta x}$

$\Rightarrow \frac{\frac{\cancel{x} - \cancel{x} - \Delta x}{x(x+\Delta x)}}{\Delta x}$

$\Leftarrow \frac{\frac{-\cancel{\Delta x}}{x(x+\Delta x)}}{\cancel{\Delta x}}$

$\Downarrow$

$\frac{-1}{x^2 + \Delta x}$ → Δx reaches 0

so $\frac{-1}{x^2} \Rightarrow \boxed{-x^{-2}}$

$f(x) = x^2$ $\quad f(x) = x^3$

$f'(x) = 2x$ $\quad f'(x) = 3x^2$

$f(x) = x^{-1}$

$f(-1)x^{-2}$

So if: $f(x) = x^n$

$$f'(x) = \frac{d}{dx} f(x) = \boxed{nx^{n-1}}$$

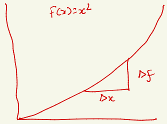
The Inverse function & it's derivative

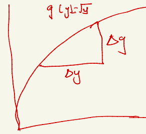
what an inverse
$0 \xrightarrow{f \text{ (goes later)}} \bullet \xrightarrow{g \text{ takes back}} 0$

so if $x \xrightarrow{f} x^2 \xrightarrow{g} x$

$g(x) = f^{-1}(x)$ $\quad g(x) = \sqrt{x}$, $f(x) = x^2$

$g(f(x)) = x$ $\quad$ for $x > 0$





The two graphs are a reflection of one another.

so $\Delta x = \Delta g$ $\quad$ so slope $\frac{\Delta g}{\Delta y} = \frac{\Delta x}{\Delta f}$
$\Delta y = \Delta f$

$\frac{\Delta f}{\Delta x} = f'(x)$ $\quad \frac{\Delta g}{\Delta y} = g'(y) = \frac{1}{f'(x)}$

for $(2,4)$ $\quad f(2) = 4$, $g(4) = 2$

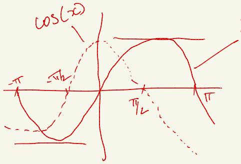
$f'(2) = 4$ $\quad g'(4) = \frac{1}{f'(4)} = \boxed{1/4}$

g' at $(1,1)$ $\quad f'(x) = 2x$ $\quad$ for $g'(y) = \frac{1}{f'(y)}$

$f'(1) = 2$ $\quad \Rightarrow 1/2$

so $\frac{(or) d(x^{-2})}{dy} \Rightarrow \frac{1}{-2} \cdot x^{-3}$

Derivative of trigonometric functions



$x = \pi/2 \quad -\pi/2 \quad 0 \quad -\pi$

$\pi = 180$ $\quad$ slope $\quad 0 \quad 0 \quad 1 \quad -1$

Now let's compare with $\quad$ ⇕ equal

$\cos(x) \to 0 \quad 0 \quad 1 \quad -1$

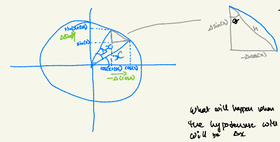
so if $f(x) = \sin(x)$

$f'(x) = \cos(x)$

Same with $\cos(x)$ $\quad f(x) = \cos(x) \to f'(x) = -\sin x$

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

How all this works



$\to \cos(\theta) = \frac{adj}{hyp} = \frac{\Delta(\sin x)}{h}$

$\Delta(\sin x) = h\cos(\theta)$ --- ①

$\to \sin(\theta) = \frac{opp}{hyp} = \frac{-\Delta(\cos x)}{h}$

$h\sin(\theta) = -\Delta(\cos x)$ --- ②

What will happen when Δx get's smaller & smaller the hypotenuse will be equal the arch which will be Δx

① & ② become:
$\Delta(\sin x) = \Delta x \cos(x)$
$-\Delta(\cos x) = \Delta x \sin(x)$

which translate to
$f(x) = \sin x \quad f'(x) = \cos x$
$f(x) = \cos x \quad f'(x) = -\sin x$

Derivative of the exponential function

$e = 2.71828182$

For n $\quad e = (1+1/n)^n$

n	1	10	100	1000	∞
$(1+1/n)^n$	2	2.594	2.705	2.717	2.718

This is almost equal to e

$p(x) = e^x$
$p'(x) = e^x$

The fact it remains the same is why it's important.

Why is it important.

Bank 1	Bank 2	Bank 3
Interest	"	"
100% every year	50% of your money every 6 months	33.3% every 4 months

→ This is best why?

	B1	B2	B3
Now	1\$	1	1
4 months	1	1	$1+1/3$
6 months	1	$1+1/2$	$1+1/3$
8 months	1	1	$(1+1/3)^2$
1 year	2	$(1+1/2)^2$	$(1+1/3)^3$

Compound the interest.

⇒ If you generalise this → 1 year
$1\ 2\ 3 \ldots\ k \ldots\ n-2\ n-1\ n$
$1/n$ interval

Interest → 1, $1 + 1/n$

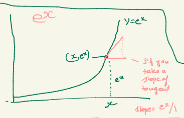
$2n \to (1+1/2)^2$ $\quad 3 \to n \Rightarrow (1+1/n)^3$

in 1 year → $(1+1/n)^n$

returns $1/n$ times your money n times a year

At the end even if they invest $1/\infty$ of your money ∞ times a year

≈ 2.71828182



Logarithm

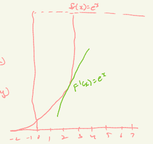
$e^? = 3 \to ? = \log(3)$

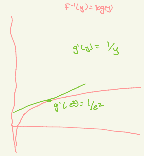
$e^? = x \to ? = \log(x)$

Natural logarithm.

$f(x) = e^x \quad f^{-1}(y) = \log(y)$

$\to e^{\log(x)} = x \quad \log(e^y) = y$





so $\frac{d}{dy} f^{-1}(y) = \frac{1}{f'(f^{-1}(y))}$

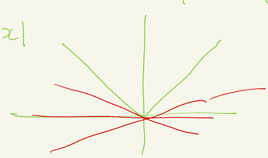
$\frac{d}{dy}(\log(y)) = \frac{1}{e^{\log(y)}}$

so $\frac{d}{dy}(\log(y)) = 1/y$

Non-Differentiable Functions

If you can draw a unique tangent on any point in a plane then the function is termed differentiable

$f(x) = |x|$

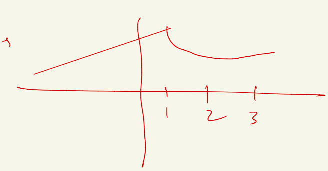


these slopes are not valid on the corner.

So. If a function forms a corner or a cusp. It does not have a derivative due to inability to draw a slope

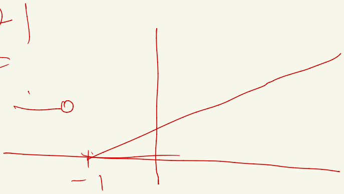
Q1

$x = -1$, $x = 0$, $x = 1$, $x = 2$ at which point does derivative of this function non differentiable,



Answer → $x = 1$

Q2



Ans → $x = -1$

→ Step function this has a jump discontinuity at $x = -1$

at $x=0$ this has a tangent that runs straight up the vertical slope. Why -> not well defined.

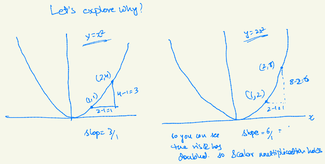
Recap: Corners | cusps, Jump discontinuity, Vertical tangents -> Non-differentiable.

Properties of derivatives.

1) multiplication by scalar -> $F=4g$ so $F'=4g'$

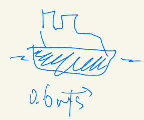
So $F=cg$, $F'=cg'$

Let's explore why?

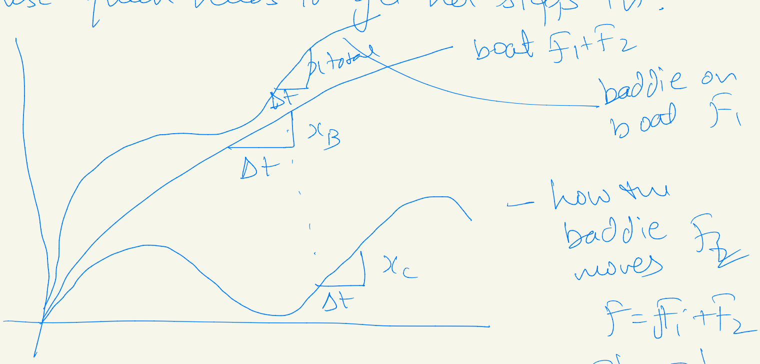


2) Sum rule.

If some gorgeous woman is on a cruise



The cruise let's say is going 0.6m/s and then the fine shyt begins walking at a speed of 0.05 m/s in the direction of the cruise, cause queen needs to get her steps in.



$x_{total} = x_B + x_C$

$$\frac{x_{total}}{\Delta t} = \frac{x_B}{\Delta t} + \frac{x_C}{\Delta t}$$

$V_{total} = V_B + V_C$

$F = F_1 + F_2$

$F' = F_1' + F_2'$?

The Product Rule

$F' = g'h + gh'$

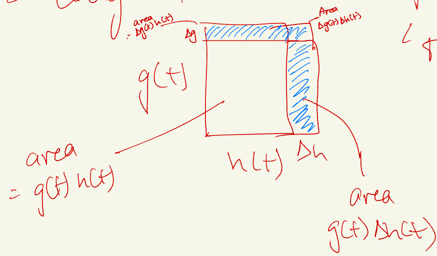
$F = gh$.

lets say a house in under construction $h(t) \to$ width of house made at time t & $g(t) \to$ length made at time t.

Area = length $\times$ width

$F(t) = g(t) \cdot h(t)$

$\to$ rate of change of area with respect to time



So the rate of change of area in a small time Δt will be

$$\frac{\Delta F(t)}{\Delta t}$$

$\Delta F(t) \to$ change in area

$$\Rightarrow \frac{\Delta F(t)}{\Delta t} = \frac{\Delta g(t) h(t) + g(t) \Delta h(t) + \Delta g(t) \Delta h(t)}{\Delta t}$$

$$\Rightarrow \frac{\Delta g(t)}{\Delta t} h(t) + g(t) \frac{\Delta h(t)}{t} + \frac{\Delta g(t) \Delta h(t)}{\Delta t}$$

Now $\Delta t \to 0$

$$g'(t) h(t) + g(t) h'(t) + 0$$

$$\to \underline{g'(t) h(t) + g(t) h'(t)}$$

Question

Q $F(x) = xe^x$

$$F'(x) = \frac{\partial(x)}{\partial x} \cdot e^x + \frac{\partial(e^x)}{dx} x$$

$$\Rightarrow e^x + xe^x$$

$$\Rightarrow \underline{e^x(x+1)}$$

The Chain Rule: $h(t), g(t) \rightarrow \frac{d}{dt}(g(h(t))) \Rightarrow \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial t} \rightarrow g'(h(t)) \cdot h'(t)$

$$\frac{\partial}{\partial t} F(g(h(t))) \rightarrow \frac{\partial F}{\partial g} \cdot \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial t} = F'(g(h(t))) * g'(h(t)) * h'(t)$$

Idea

cold

mountain hot

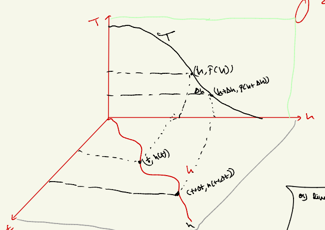
(1) As you drive up the temp changes w.r.t height $\rightarrow \frac{\partial T}{\partial h}$ - temp, - height

(2) But also the height changes as you drive up so height changes w.r.t time t, time $\frac{\partial h}{\partial t}$

(3) As you drive up the temp changes with time passing so $\frac{\partial T}{\partial t}$

So what the chain rule says is you can find (3) by using (1) & (2)

$$\frac{\partial T}{\partial t} = \frac{\partial T}{\partial h} \cdot \frac{\partial h}{\partial t}$$



as time changes, height changes, Temperature changes.

$\Delta t \rightarrow \Delta h \rightarrow \Delta T$

$$\frac{\Delta T}{\Delta t} = \frac{\Delta T}{\Delta h} \cdot \frac{\Delta h}{\Delta t}$$

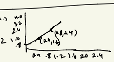
$$\frac{\partial T}{\partial t} = \frac{\partial T}{\partial h} \frac{\partial h}{\partial t}$$

Q) $F(x) = e^{2x}$ $F'(x) = ?$

$e^{2x} \cdot 2 \Rightarrow 2e^{2x}$

WEEK 1 Practice Assignment

Q) line 1 slope < line 2 slope. Q 2) mini graph $\rightarrow$



$$\text{slope} = \frac{\text{rise}}{\text{run}} \Rightarrow \frac{2.4 - 1.6}{0.8 - 0.4} \Rightarrow \frac{0.8}{0.2} = 2$$

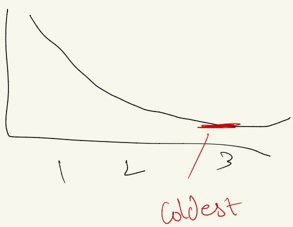
Q 4) $y_1 = ax + b$ $y_2 = cx + d$ $a, b, c, d \in R$

slope $y_1 =$

Why Are derivatives used for ML.

In ML we find an ML model that best fit's your Dataset and in order to find this model you calculate a loss function that tells you how far you are from the ideal model. & when you minimize this function you get the best model.

At the maximum or min. the slope is zero.

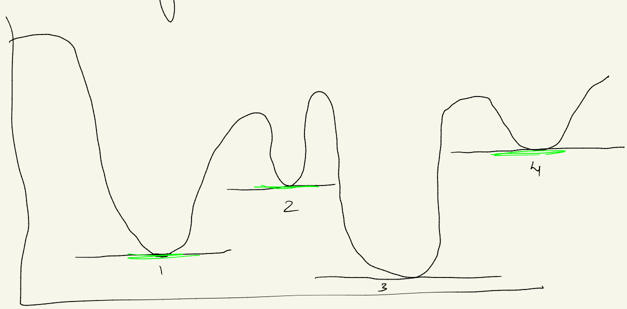


you move from left to right in a sauna you see if hot or cold.

At the coldest point, when you move left or right the ~~temp~~ increases.

[I'm guessing the actual solution is where problem is correct. Any right slope changes cause you move away from the ideal point]

Not always true that slope = 0 means max or min.



However you are able to narrow down the results to look for.

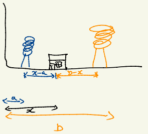
local minimum

maximum | global minimum.

Cost Function Motivation

Let's say one power hub, you put house next to powerline $x=0=0$

Now let's say 2. blue & orange.



So the cost of connecting is distance^2

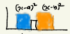
blue → $(x-a)^2$

orange → $(b-x)^2 \to (x-b)^2$

Total cost of connecting to both powerlines

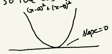
$(x-a)^2 + (x-b)^2$

so let's visualise the cost as the area of square so



So these two squares

So the cost function → $(x-a)^2 + (x-b)^2$



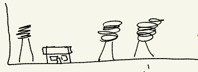
$$\frac{d}{dx}\left[(x-a)^2 + (x-b)^2\right] = 0$$

$$2(x-a) + 2(x-b) = 0$$

$$x-a + x-b = 0$$

$$2x = a+b$$

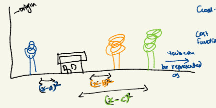
$$x = \frac{a+b}{2} \to \boxed{\text{half way distance b/w both.}}$$



connecting to power grids is expensive tower

cost of connecting to power line is x^2. The goal minimise the cost of connecting to all 3 power lines

Now more:



Goal → minimize of cost of collecting to the power lines

Cost function → $(x-a)^2 + (x-b)^2 + (x-c)^2$

minimize this.

for slope $=0$ $\quad \frac{d}{dx}\left[(x-a)^2 + (x-b)^2 + (x-c)^2\right] = 0$

$$2(x-a) + 2(x-b) + 2(x-c) = 0$$

$$2x - 2a + 2x - 2b + 2x - 2c = 0$$

$$6x = 2a + 2b + 2c$$

$$x = \frac{\cancel{2}(a+b+c)}{\cancel{6}_3} \to \boxed{\frac{a+b+c}{3}}$$

The square loss

So if you were to have n power lines: minimize $(x-a_1)^2 + (x-a_2)^2 + \dots + (x-a_n)^2$

solution $= x = \dfrac{a_1 + a_2 + \dots + a_n}{n}$

The log loss → let's say 10 times coin flipped. you want 7H 3T

Now let's say H → 'p' probability T → 1-p

so 7H 3T → $p^7(1-p)^3 = g(p)$

Now we use calculus to maximise $g(p)$

$$\frac{dg}{dp} = \frac{d}{dp}\left(p^7(1-p^3)\right) = \frac{d(p^7)}{dp}(1-p^3) + p^7 \frac{d((1-p)^3)}{dp}$$ — chain rule.

$$= 7p^6(1-p)^3 + p^7 3(1-p)^2(-1)$$

$$= p^6(1-p)^2\left[7(1-p) - 3p\right]$$

$$= p^6(1-p)^2(7-10p) = 0$$

⇒ so for a product of 3 things to be zero 1 must be zero

for $p=0$, $p=1$, $p=0.7$ $\quad \boxed{p=0.7 \text{ will work}}$

If $p=0$ it will always lands on tails

$p=1$ " " " " heads

How to make this easier

$\log(g(p))$ → because if $g(p)$ is maximal so is the logarithmic.

so maximizing the logarithm will maximize $g(p)$

$$\log(g(p)) = \log\left(p^7(1-p)^3\right) = \log(p^7) + \log\left((1-p)^3\right)$$

$$= 7\log(p) + 3\log(1-p) = G(p)$$

Now take derivative

$$\frac{dG(p)}{dp} = \frac{d}{dp}\left(7\log(p) + 3\log(1-p)\right) = 7\frac{1}{p} + 3\frac{1}{1-p}(-1) = \frac{7(1-p) - 3p}{p(1-p)} = 0$$

$p = 0.7 \leftarrow 7(1-p) - 3p = 0$

Now we get same solution in a simpler way. logarithms of probability are very very common in ML. what we need is $-G(p) \to$ log loss

why do we take $-G(p)$

$\log(p)$ is a -ve no. when p is b/w 0 & 1

Relationship with ML

H H H H H H H T T T

$\underbrace{H\ H\ H\ H\ H\ H\ H}_{7}\ \underbrace{T\ T\ T}_{3}$

We found model that get's 7 H 3 T how? by minimizing the log loss & obtained that optimal p is 0.7

Why log?

1) Derivatives of products is hard, derivatives of sums is easy.

eg)
$$F(p) = p^6(1-p)^2(3-p)^9(p-4)^{13}(10-p)^{500}.$$

direct very hard.

$$\frac{\partial}{\partial p}\log(F) = \text{easier}$$

2) Product of tiny things is tiny.

small no.s x small no.s = even smaller no.s

log(small no.) = - large no.s → These can computed by computer easily.

— Week 1 theory End —