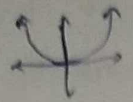
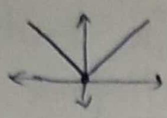
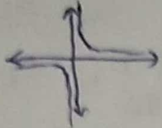
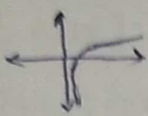
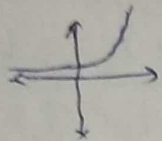
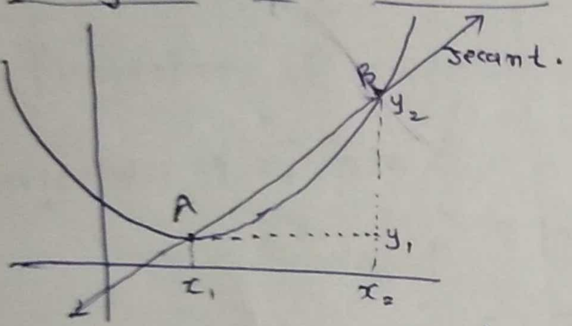
* Calculus *

* Limits range and domain of functions :-

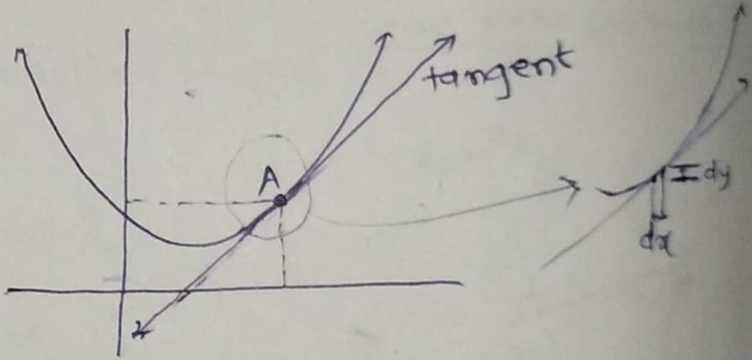
	function	graph	Domain	Range
①	x^2		R	R_0^+
②	$\lvert x \rvert$		R	R_0^+
③	$1/x$		$R-\{0\}$	$R-\{0\}$
④	$\ln(x)$		R^+	R
⑤	e^x		R	R^+

* Tangent vs secant



secant gives rate of change of function in range

slope (change) = $\dfrac{y_2 - y_1}{x_2 - x_1}$



when $(x_2 - x_1) \to 0$, i.e. $dx \to 0$
line touches the function at one and only point.

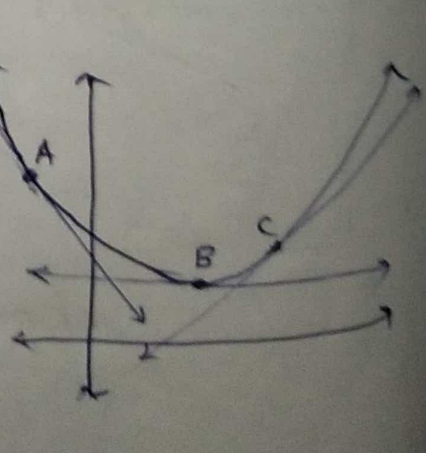
Slope = $\dfrac{dy}{dx}$ — differentiation @ point A.

* Derivation / differentiation :-

"Rate of change of function at point x." It is also defined using limit as,

$$f'(x) = \frac{df}{dx} = \lim_{dx \to 0} \frac{f(x+dx) - f(x)}{dx}$$

point	$\left.\frac{df}{dx}\right\rvert_{point}$	interpretation
A	<0	f is decreasing
B	$=0$	f is optimum
C	>0	f is increasing



* Continuity and differentiability :-

☆ Function is continuous at point a if,

1) f(a) is defined

2) $\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = f(a)$

☆ Function is differentiable at point a if,

$$f'(a) = \left.\frac{df}{dx}\right|_{x=a} = \lim_{dx \to 0} \frac{f(x+dx) - f(x)}{dx}$$ is defined at $\underline{\underline{x=a}}$

$dx \to 0^+$ and $dx \to 0^-$

EXAMPLE: ① $f(x) = \frac{1}{x}$ is not defined at $x=0$. So it is not continuous

$\frac{d}{dx}\left(\frac{1}{x}\right)$ is not defined at $x=0$. So it is not differentiable.

② $|x| \Rightarrow$ is defined at $x=0$, $|x| = |0| = 0$. It is continuos.

But $\frac{d}{dx}|x|$ is not defined at $x=0$. So it is not differentiable @ $x=0$.

We can observe $|x|$ is not smooth @ $x=0$ ✓

* RULES of differentiation :-

① $(f \pm g)' = f' \pm g'$ $\quad \frac{d}{dx}(x^2 + 2x - e^x) = 2x + 2 - e^x$

② $(f \cdot g)' = f'g + g'f$ $\quad \frac{d}{dx}(x^3 \cdot e^x) = 3x^2 \cdot e^x + x^3 \cdot e^x$

OR $f'g + fg'$

③ $\left(\frac{f}{g}\right)' = \frac{f' \cdot g - g' \cdot f}{g^2}$ $\quad \frac{d}{dx}\left(\frac{\sin x}{x^2}\right) = \frac{\cos x \cdot x^2 - \sin x \cdot 2x}{x^4}$

④ $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$ $\quad \frac{d}{dx}(\sin(x^2)) = \underset{f'}{\cos(x^2)} \cdot \underset{g'}{2x}$

$(f \circ g \circ h)'(x) = f'(g(h(x))) \cdot g'(h(x)) \cdot h'(x)$

$$\frac{d}{dx} = e^{\sin(x^2)} = \underbrace{e^{\sin(x^2)}}_{f'} \cdot \underbrace{\cos(x^2)}_{g'} \cdot \underbrace{2x}_{h'}$$

Variation.2 $\quad y = f(u)$ and $u = f(x)$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$y = 2u^2$. $u = x^3$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$$\frac{dy}{dx} = \frac{d}{du}(u^2) \cdot \frac{d}{dx}(x^3)$$

$= 2 \cdot u \cdot 3x^2$

$= 2u(x^2) \cdot 3x^2$

$= 6x^5$

$y = 2u^2 = 2(x^3)^2 = 2x^6$ $\quad \therefore 6x^5$

** OPTIMA OF A FUNCTION

* PARTIAL derrivative :-

$f(x,y) = x \cdot y$

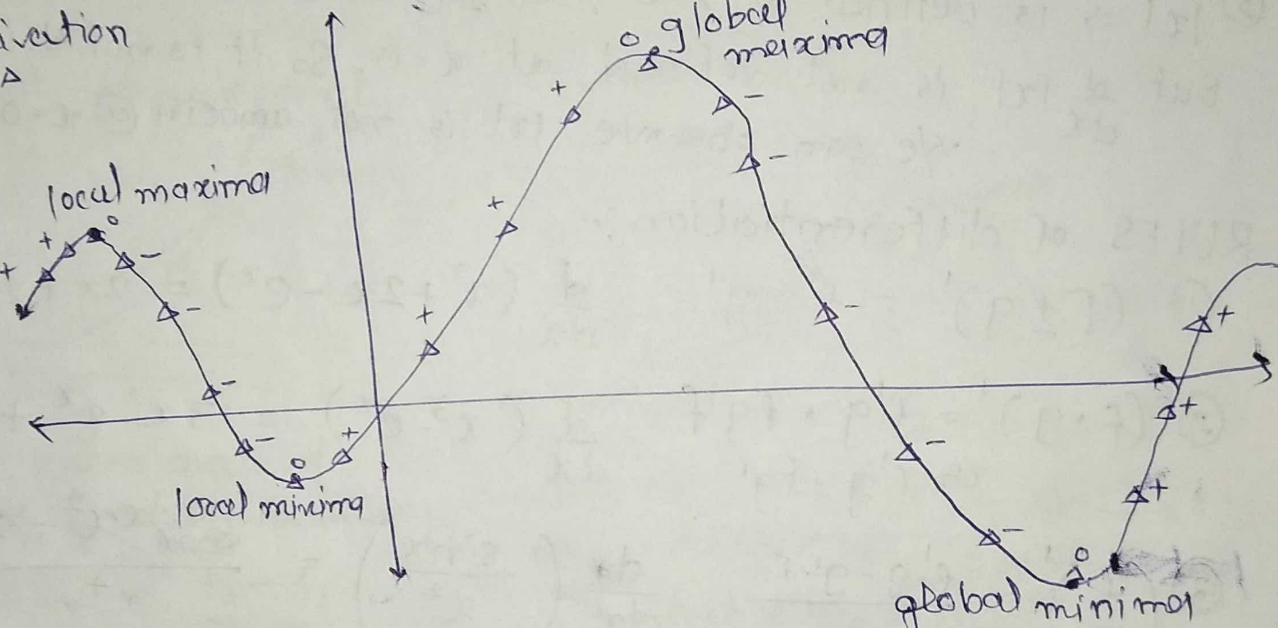
In case of multivariable function, we can differentiate function with respect to one variable which is defined as partial derrivative.

$\frac{\partial f}{\partial x} = y \cdot \frac{\partial x}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \cdot \frac{\partial y}{\partial y} = x$

Ex. $f(x,y) = x^3 + y^2$

$\frac{\partial f}{\partial x} = 3x^2 + 0 = 3x^2$ and $\frac{\partial f}{\partial y} = 0 + 2y = 2y$

=> Minima and Maxima



* $\xrightarrow{+++} 0 \xrightarrow{---}$ Maxima

$\xrightarrow{---} 0 \xrightarrow{+++}$ Minima

=> at optima $\frac{df}{dx} = f'(x) = 0$,

By evaluating $f'(x) = 0$, Roots $= C_1, C_2, \dots C_n$

For each root calculate

$$f''(C_i) \begin{cases} < 0, & \text{Maxima at } C_i \\ > 0, & \text{Minima at } C_i \\ = 0, & \text{CND. Investigate further.} \end{cases}$$

Intuition

https://www.geogebra.org/calculator/pm7s54vc

$f'(x) = 0$, f is optimum, $f''(x) > 0$ means $f'(x)$ is increasing

$f''(x) < 0$ means $f'(x)$ is decreasing.

⇒ OPTIMA of Multivariable function

$z = f(x, y) = x^2 - y^2$

here $f_x = \frac{\partial f}{\partial x}$, $f_{xx} = \frac{\partial^2 f}{\partial x^2}$, $f_y = \frac{\partial f}{\partial y}$, $f_{yy} = \frac{\partial^2 f}{\partial y^2}$

$f_{xy} = f_{yx} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$

1) Evaluate $f_x = 0$ and $f_y = 0$, Roots = $(a_1, b_1), (a_2, b_2), \dots (a_n, b_n)$

2) For each root calculate $D_{@(a_i, b_i)} = f_{xx} \cdot f_{yy} - f_{xy}^2$

case I $D > 0$, f is optimum

- case I.I : $f_{xx} < 0$, $f_{yy} < 0$ ___ f is maxima
- case I.II : $f_{xx} > 0$, $f_{yy} > 0$ ___ f is minima

case II $D < 0$, saddle point at (a_i, b_i)

case III $D = 0$, CND.

---

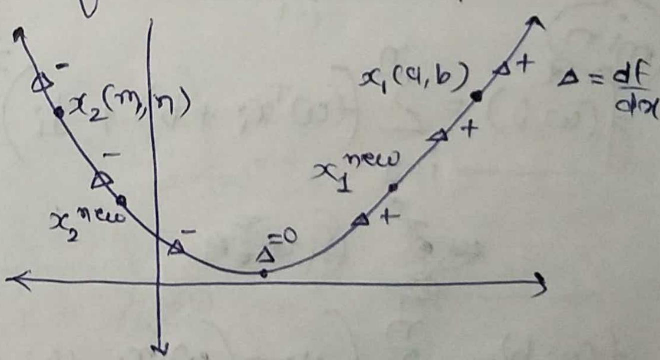
* Above methods are not good in higher order functions. We use gradient Descent method to find minima.

$f(x) = x^6 + x^5 + 4x^2$ $\quad f'(x) = \underbrace{6x^5 + 5x^4 + 8x = 0}_{\text{Very difficult to find roots.}}$

How gradient descent works to find Minima?

→ If Δ is positive, f is increasing, we are climbing the hill. To go back to valley we need to go in opposite direction.

→ If Δ is negative, f is decreasing, we are walking down to valley. To go down to valley we need to go in same direction.



CASE 1 ($\lambda > 0$) we are at $(a, b) = x_1$

Δ is negative

$x_2^{new} = x_2^{old} + \lambda \cdot (-\Delta)$

$\Delta < 0, \therefore \lambda \cdot (-\Delta) > 0$

$\therefore \boxed{x_2^{new} > x_2^{old}}$ → we walk towards valley

CASE 2 ($\lambda > 0$) we are at $x_2 = (m, n)$

Δ is positive

$x_2^{new} = x_2^{old} + \lambda \cdot (-\Delta)$

$\Delta > 0 \therefore \lambda \cdot (-\Delta) < 0$

$\therefore \boxed{x_2^{new} < x_2^{old}}$ → we walk towards valley

* So we can find minima by repeating,

$$x^{new} = x^{old} + \lambda \cdot (- f'_{x^{old}})$$

* For multivariable function,

$$z = f(x, y)$$

$$x^{new} = x^{old} + \lambda(- f_x(x^{old}, y^{old})) \qquad f_x = \frac{\partial f}{\partial x}$$

$$y^{new} = y^{old} + \lambda(- f_y(x^{old}, y^{old})) \qquad f_y = \frac{\partial f}{\partial y}$$

* Regression problem and solve using G.D.

i^{th} data point

$x_i = [x_{i1}, x_{i2}, \dots x_{id}]$

i^{th} output value y_i

Error = $y_i - y_{pred_i}$

$x_i \downarrow$	x_{i1}	x_{i2}	...	x_{id}	y_i
x_1	x_{11}	x_{12}	...	x_{1d}	y_1
x_2	x_{21}	x_{22}	...	x_{2d}	y_2
x_3	x_{31}	x_{32}	...	x_{3d}	y_3
⋮	⋮				⋮
x_n	x_{n1}	x_{n2}	...	x_{nd}	y_n

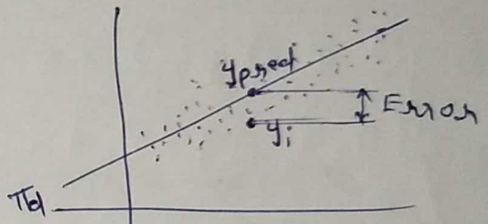
n datapoints; features, dimension; output value

MODEL REGR $\rightarrow$

y_{pred}
y_{pred1}
y_{pred2}
y_{pred3}
⋮
y_{predn}

Total Error $= \sum_{i=1}^{n} (y_i - y_{pred})$

Now our Job is to find hyperplane $\pi_d = w^T x_i + b$ S.T. sum of squares of Error is minimum.



Error $E_i = y_i - y_{pred}$

$E_i = y_i - (w^T x_i + b)$

min $E_{Total} = f(w,b) = \sum_{i=1}^{n} (y_i - (w^T x_i + b))^2$

$f(w,b) = \sum_{i=1}^{n} (y_i - (w_1 x_{i1} + w_2 x_{i2} + \dots + w_d x_{id} + b))^2$

To minimize $f(w,b)$ or $f(w_1, w_2, w_3, \dots, w_d, b)$ we need to find $\frac{\partial f}{\partial w_j} = 0$ $j = 1$ to d and $\frac{\partial f}{\partial b} = 0 \implies$ This is tough

So, To, use gradiant descent

$$w_j^{new} = w_j^{old} + \lambda \left(-\frac{\partial f}{\partial w_j}\right)_{(w^{old}, b^{old})} \quad \text{--- (1)}$$

$$b^{new} = b^{old} + \lambda \left(-\frac{\partial f}{\partial b}\right)_{(w^{old}, b^{old})} \quad \text{--- (2)}$$

We need to compute $\frac{\partial f}{\partial w_j}$ and $\frac{\partial f}{\partial b}$ for G.D. eq^n.

$$f(w,b) = \sum_{i=1}^{n} (y_i - (w^T x_i + b))^2$$

$$\therefore \frac{\partial f}{\partial w_j} = \sum_{i=1}^{n} 2(y_i - (w^T x_i + b)) \cdot \frac{\partial f}{\partial w_j}\left(y_i - (w_1 x_{i1} + w_2 x_{i2} + \dots + w_d \cdot x_{id} + b)\right)$$

$$\boxed{\frac{\partial f}{\partial w_j} = \sum_{i=1}^{n} -2(x_{ij})(y_i - (w^T x_i + b))}$$

similarly

$$\frac{\partial f}{\partial b} = \sum_{i=1}^{n} 2\cdot\left(y_i - (w^T x_i + b)\right)\cdot \frac{\partial f}{\partial b}\left(\overset{0}{y_i} - \overset{0}{w^T x_i} - \overset{-1}{b}\right)$$

$$\boxed{\frac{\partial f}{\partial b} = \sum_{i=1}^{n} (-2)\cdot\left(y_i - (w^T x_i + b)\right)}$$

By repeating eq^n 1 and 2, we find optimum w^*, b^* such that $E_{Total} = F(w, b)$ is minimum, i.e. Error is minimum and y_{pred} is near to y_i.