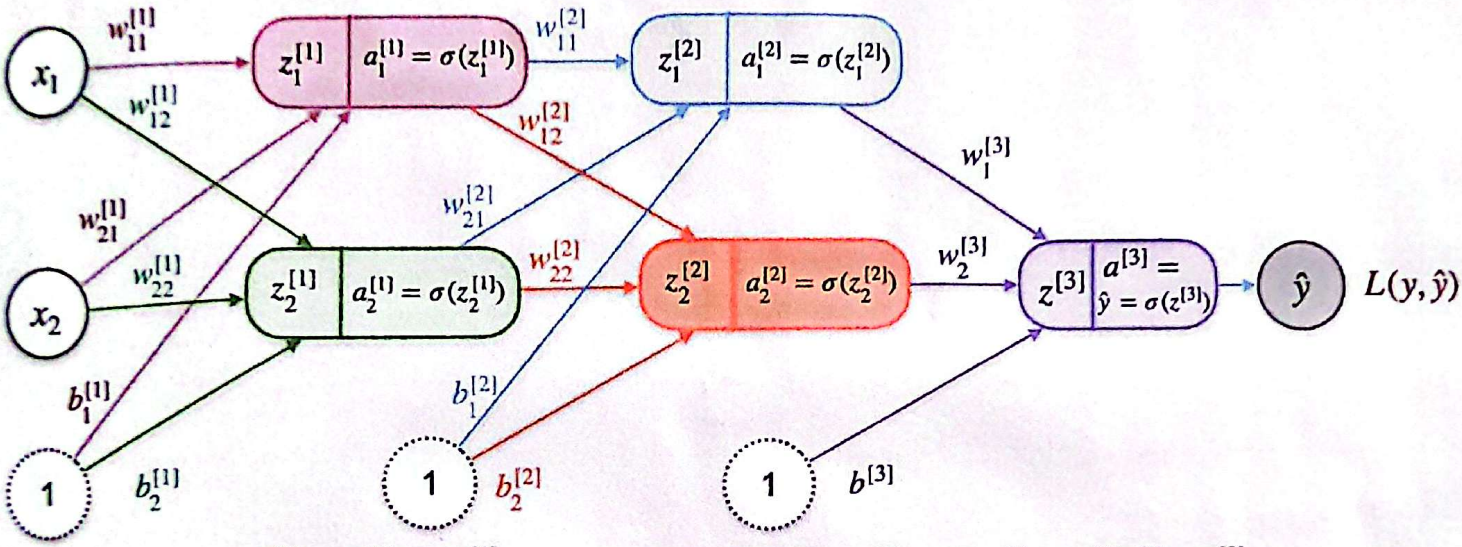
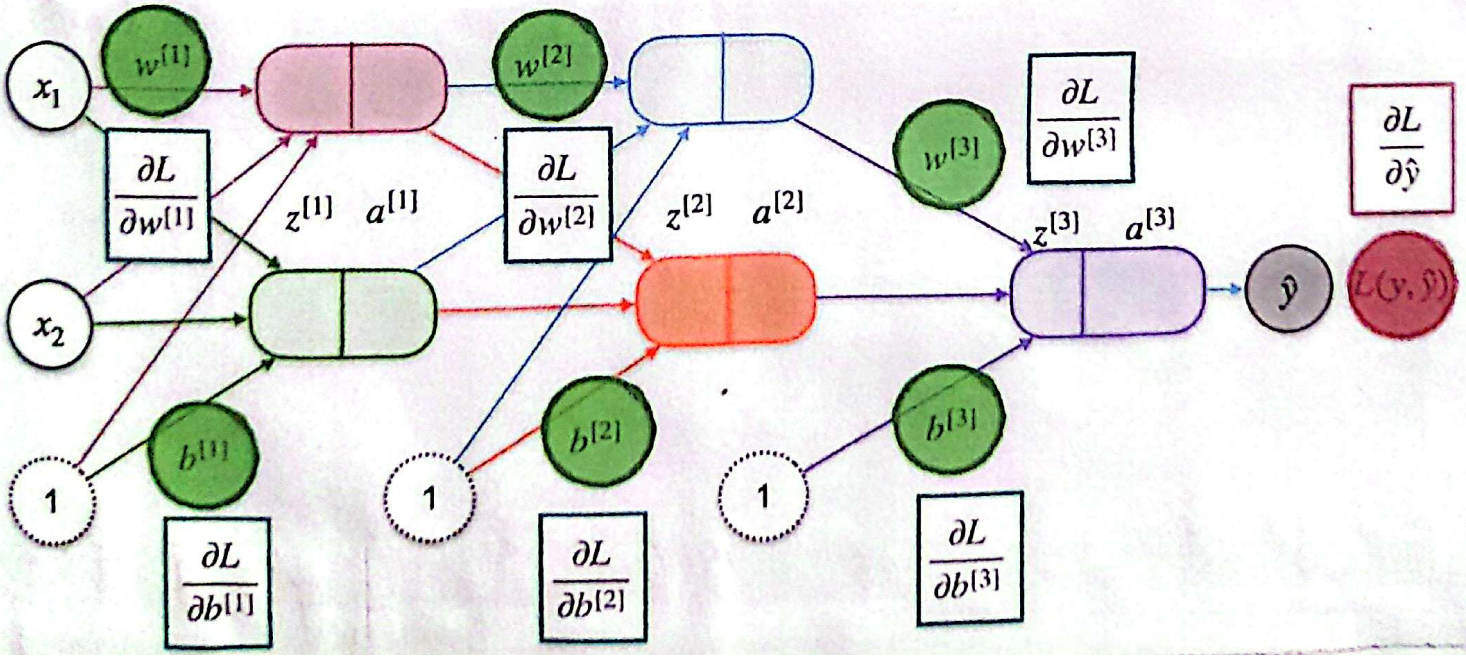
Back Propagation Introduction

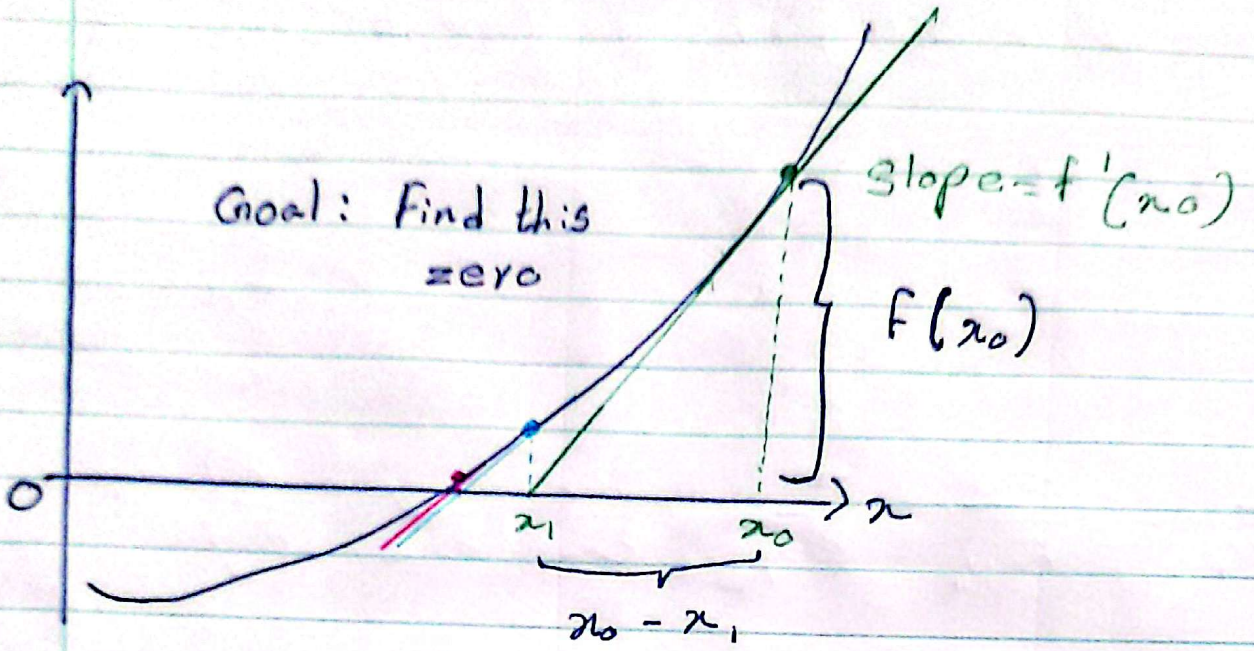


$$z^{[1]} = w^{[1]}x + b^{[1]} \qquad z^{[2]} = w^{[2]}a^{[1]} + b^{[2]} \qquad z^{[3]} = w^{[3]}a^{[2]} + b^{[3]}$$

$$L(y, \hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$



9] Newton's Method



$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

$$\frac{F(x_0)}{f'(x_0)} = x_0 - x_1$$

$$\boxed{x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}}$$

Newton's Method	NM for optimization
Goal: find a zero of $f(x)$	Goal: minimize $g(x) \rightarrow$ find zeros of $g'(x)$
	$f(x) \rightarrow g'(x)$
	$f'(x) \rightarrow (g'(x))'$
1] Start with some x_0	1] Start with some x_0
2] Update: $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$	2] Update: $x_{k+1} = x_k - \frac{g'(x_k)}{(g'(x_k))'}$
3] Repeat 2 until you find the root	3] Repeat 2 until you find the candidate for

$g(x) = e^x - \log(x)$ $\quad g'(x) = e^x - \frac{1}{x}$

Minimum $\Rightarrow 0.567$

$x_0 = 0.05$

$$(g'(x))' = e^x + \frac{1}{x^2}$$

$$x_1 = x_0 - \frac{g'(x_0)}{(g'(x_0))'}$$

$$= 0.05 - \frac{\left(e^{0.05} - \frac{1}{0.05}\right)}{e^{0.05} + \frac{1}{0.05^2}} = 0.097$$

$x_1 = 0.097$

Similarly $\longrightarrow$ $x_2 = 0.183$

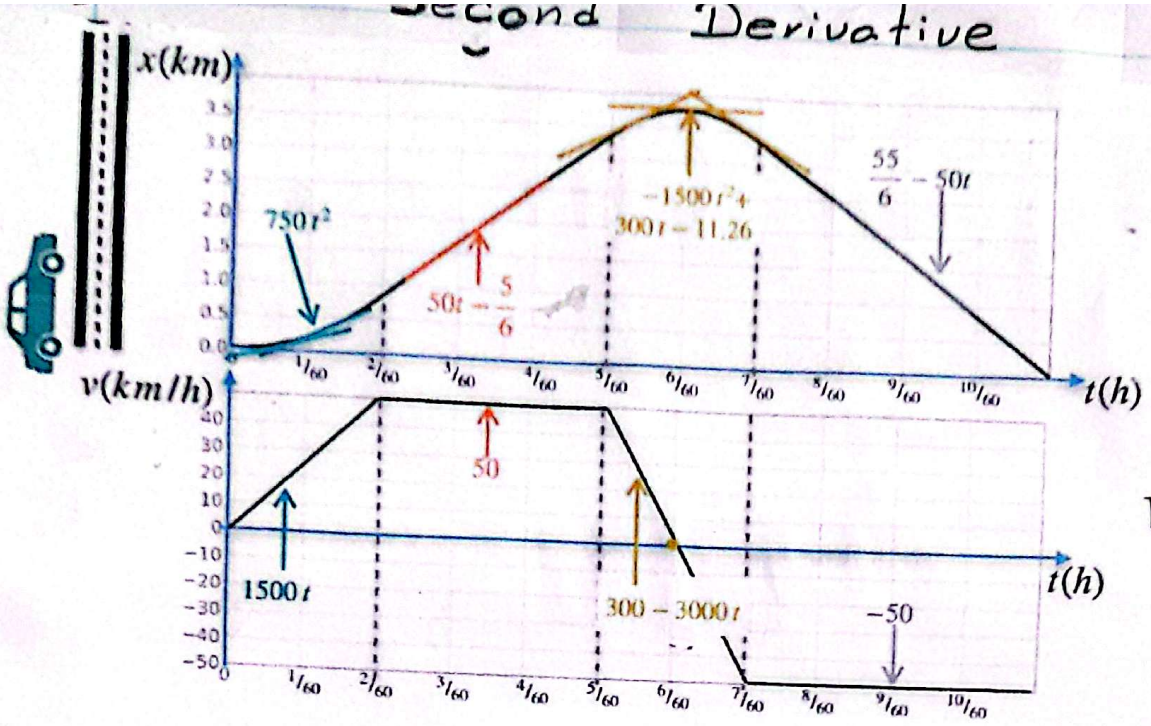
$x_3 = 0.320$

$x_4 = 0.477$

$x_5 = 0.558$

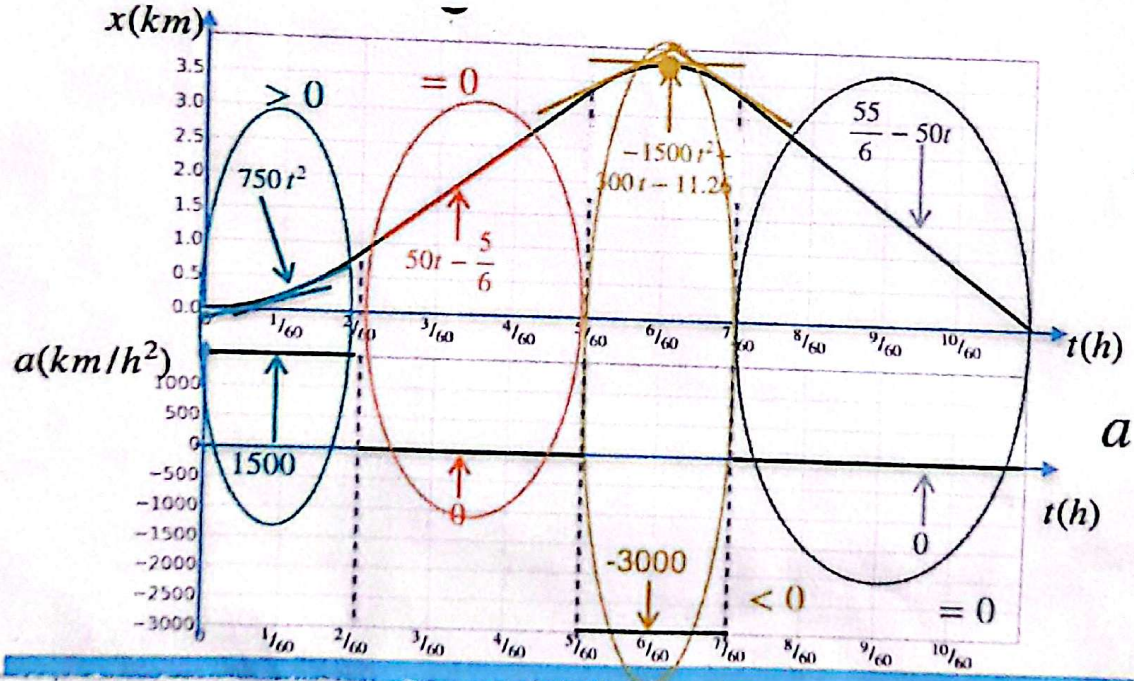
$x_6 = 0.567$ (reach the minimum)

Second Derivative



x Distance

v Velocity $\dfrac{dx}{dt}$



x Distance

Second derivative tells us about the curvature

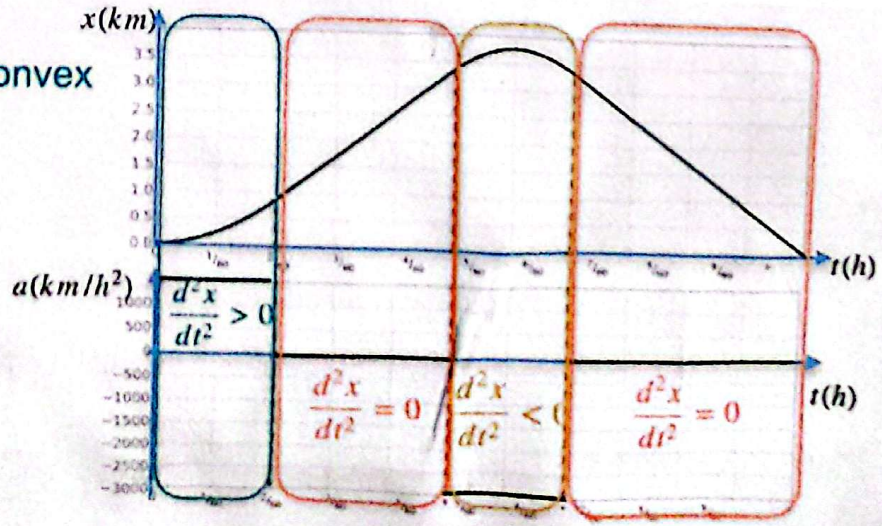
a Acceleration $\dfrac{d^2x}{dt^2}$

$\dfrac{d^2x}{dt^2} > 0$ Concave up or convex

$\dfrac{d^2x}{dt^2} < 0$ Concave down

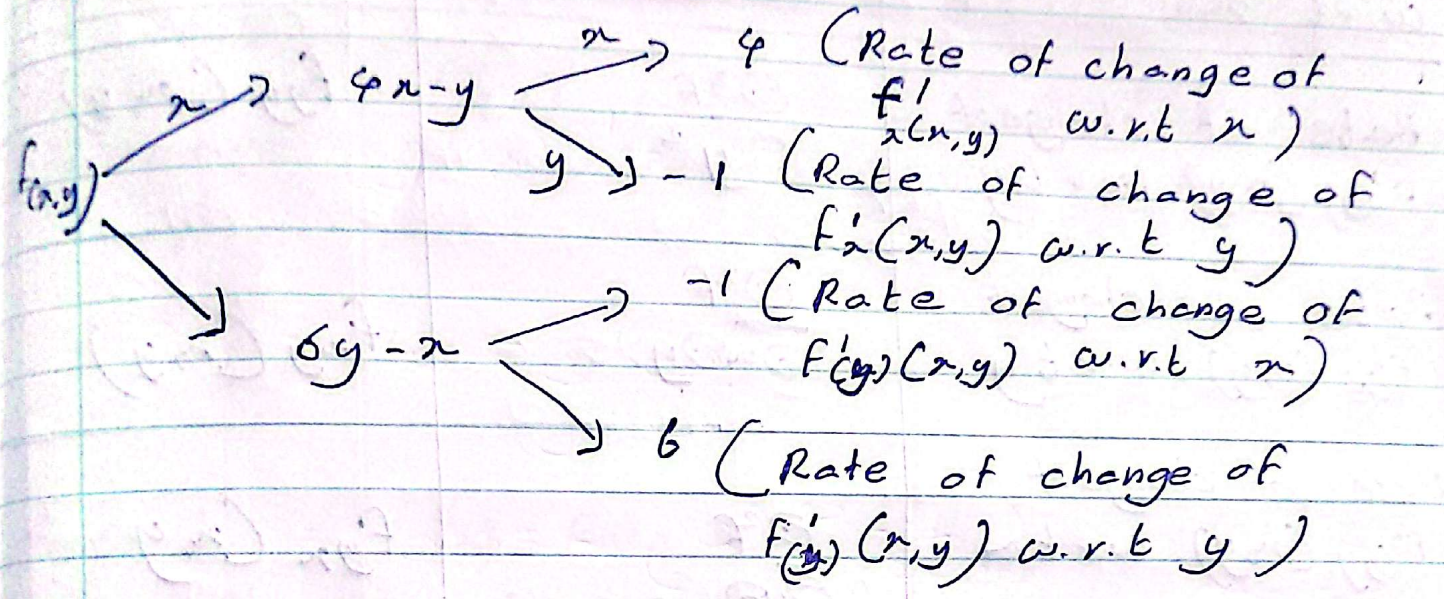
$\dfrac{d^2x}{dt^2} = 0$ Need more information

Curvature



11] Hessian

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



Rate of change of $f'_x(x,y)$ w.r.t x

Rate of change of $f'_y(x,y)$ w.r.t y

} Change in the change in the function w.r.t tiny changes in x and y

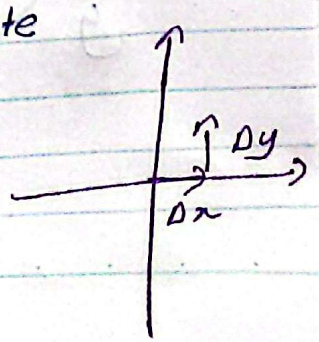
Same idea with one Variable

Rate of change of $f_x(x,y)$ w.r.t y

Rate of change of $f_y(x,y)$ w.r.t x

} 1] Change in the slope along one coordinate axis w.r.t tiny changes along an orthogonal coordinate axis

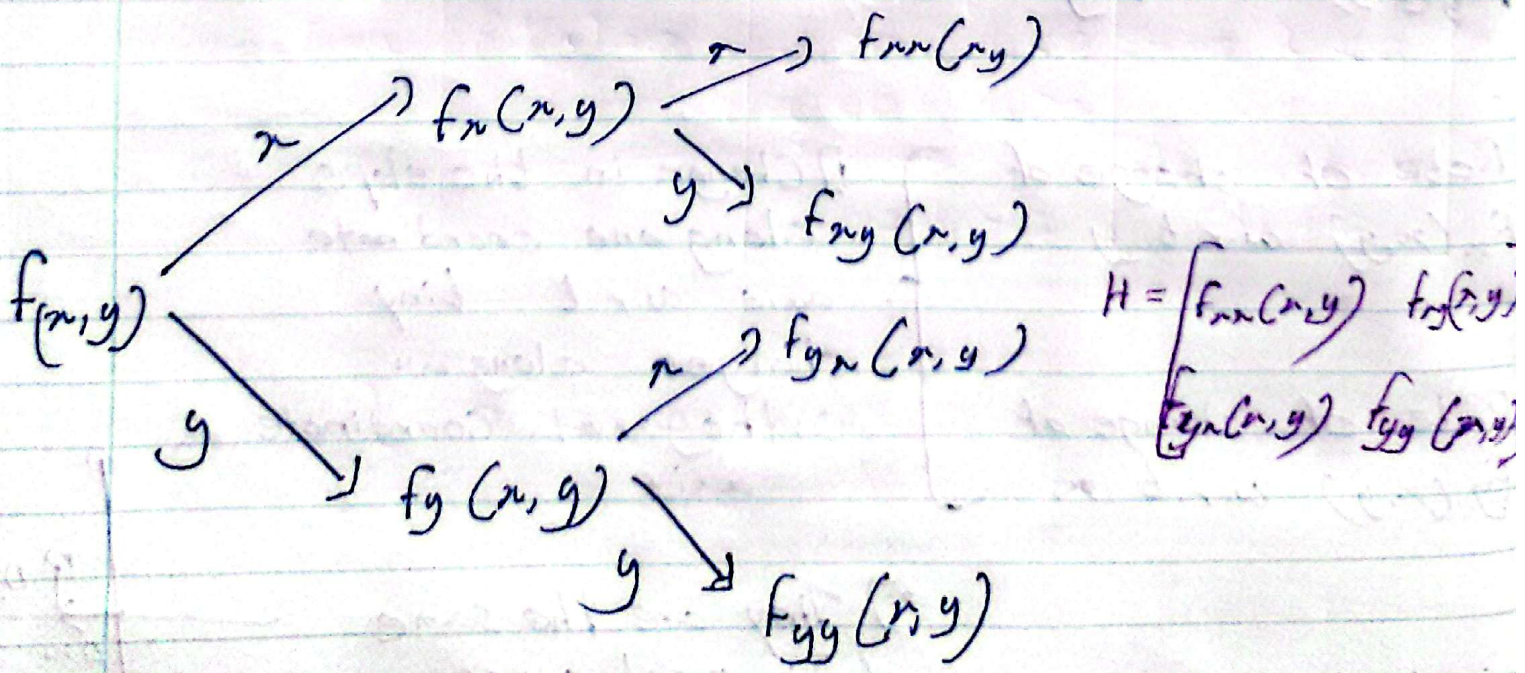
2] They are the same (In most cases



	Leibniz's Notation	Lagrange's Nota...
* Rate of change of $f'(x)(x,y)$ w.r.t x	$\frac{\partial^2 f}{\partial x^2}$	$f_{xx}(x,y)$
Rate of change of $f'(y)(x,y)$ w.r.t y	$\frac{\partial^2 f}{\partial y^2}$	$f_{yy}(x,y)$
Rate of change of $f'(x)(x,y)$ w.r.t y	$\frac{\partial^2 f}{\partial x \partial y}$	$f_{xy}(x,y)$
Rate of change of $f'(y)(x,y)$ w.r.t x	$\frac{\partial^2 f}{\partial y \partial x}$	$f_{yx}(x,y)$

Hessia matrix

$$H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$



	1 Variable	2 variables
Function	$f(x)$	$f(x,y)$
first derivative	$f'(x)$ Rate of Change of $f(x)$	$F_x(x,y)$ R.O.F w.r.t x $F_y(x,y)$ R.O.F w.r.t y $\nabla F = \begin{bmatrix} f_x(x,y) \\ f_y(x,y) \end{bmatrix}$
Second derivative	$f''(x)$ Rate of change of the rate of Change of $f(x)$	$H_{(x,y)} = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$

11] Hessians and Concavity

	1 Variable $f(x)$	2 Variable $f(x,y)$	More Variables $f(x_1, x_2 \dots x_n)$
(Local) minima	Happy face $f''(x) > 0$	Concave up Upper paraboloid $\lambda_1 > 0$ & $\lambda_2 > 0$	All $\lambda_i > 0$
(Local) maxima	Sad face $f''(x) < 0$	Concave down Down paraboloid $\lambda_1 < 0$ & $\lambda_2 < 0$	All $\lambda_i < 0$
Need more information	$f''(x) = 0$	Saddle point $\lambda_1 > 0$ & $\lambda_2 < 0$ $\lambda_1 < 0$ & $\lambda_2 > 0$ or Some $\lambda_i = 0$	Some $\lambda_i > 0$ Some $\lambda_j < 0$ or At least one $\lambda_j = 0$

λ = Eigen Values