UNIT-I

Introduction to Machine Learning

Machine Learning (ML):

ML is a Branch of Artificial Intelligence (AI) that enables computers to self-learn from training data and improve over time without being explicitly programmed

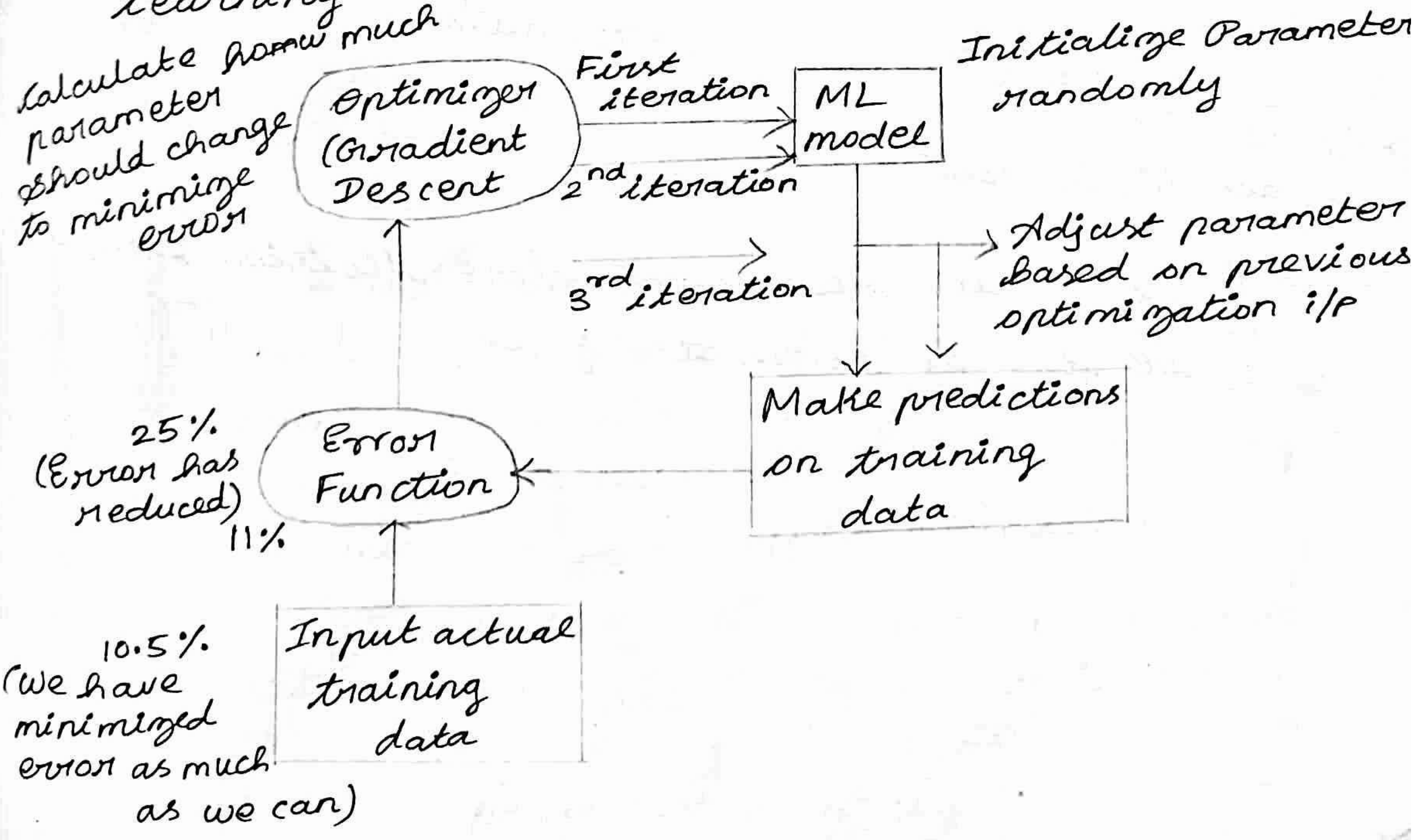
Introduction

* Optimization is one of the core components of machine learning

* The essence of most machine learning algorithms is to build an optimization model and learn the parameters in the OF from the given data

Role of Optimization.

Optimization is a Supervised machine learning



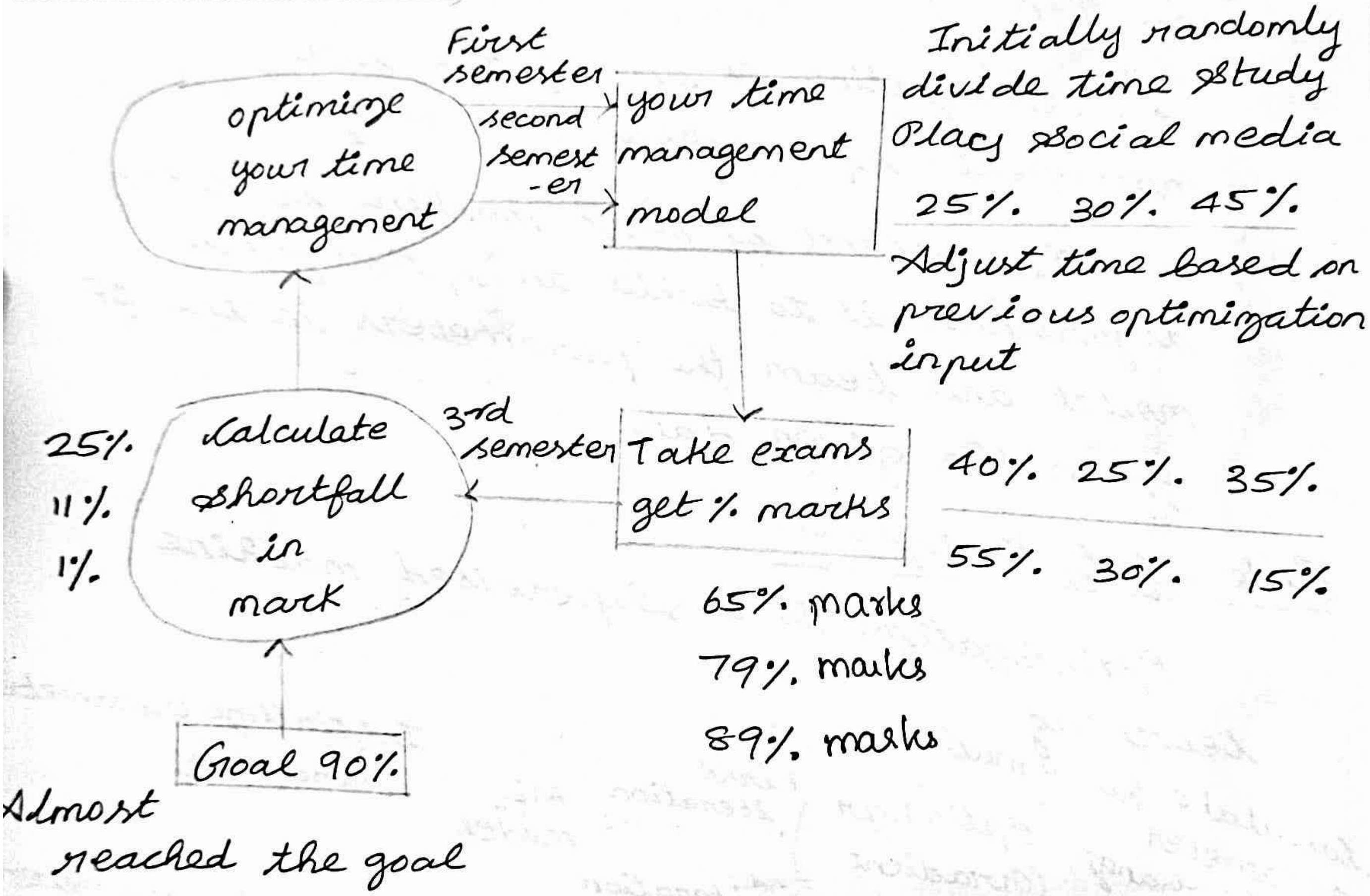
(2) * Error Functions are known as loss functions

* The optimizer used in supervised learning is GD.

* The iterations are known as epoch (It may vary from few iteration to thousands)

* The model thus obtained is a trained mode. This model can be used to make predictions on unseen test data to verify the accuracy of the model

Real life Example:



Calculate how much time division should be done to reach the goal

(3) * Here we have a model that initially set random values for its parameter. These parameters helps to build a function (eg.)

$$\boxed{y = w_0 x_0 + w_1 x_1 + w_2 x_2}$$

where x_0 x_1 x_2 are features (think, study, play, social media)

w_0 w_1 w_2 are weights (time given to play, social media

Y- o/put or prediction (exam score)

This function is used to make prediction on training data set

* This prediction is then compared to the actual result of training set

* Both predicted and actual o/p is then sent to a error function. This error functions calculates the offset or error between the predicted and actual output

* This error is then sent to an optimizer The optimizer calculates that how much the initial values of the weights should be changed so that the error is reduced further and we move towards expected o/p

* The weights of the model are adjusted according for next iterations and again prediction are made on training set, the error is calculated and optimizer again recommends for weight adjustment.

* These iterations should go on until not much changes in the error is reached or we have reached the desired goals in terms of prediction accuracy. At this point the iteration should be stopped.

Mathematical Optimization

* In optimization we will deal with minimize/maximize some function subject to constraints

$$\boxed{\begin{array}{l} \text{minimize } f_0(x) \\ \quad \text{s.t. } f_i(x) \leq b_i, \; i = 1 \ldots m \end{array}}$$

* The function we wish to maximize/minimize is the OF

* Constraints are inequality constraint function

* Those constraints have bounds which are constants in the constraint function

Linear Optimization

(objective function)

* If the OF and CF are linear
(constraint function)

Convex Optimization

* If the OF and CF are convex
* Any linear program is a convex optimization problem

Methodology

* The variable x in the above equation represents the choice made

The constraint $f_i(x) \leq b_i$ represent limitation ~~limination~~ on the possible choices

The OF $f_0(x)$ represent the cost of choosing x.

So a solution to the optimization problem means finding a choice that has a means finding a choice that has a minimum cost among all the choices the meet the requirements

Applications:-

1) Portfolio Optimization:-

* Portfolio - is a collection of financial investment like stocks, bonds, short term investments etc

* The problem is to seek the best way to invest some capital in a set of assets

(ie) To build a portfolio in such a way that you maximize potential returns. from investments while still not exceeding the amount of risk you are willing to carry

* Variable 'x' represent the investment made

* Constraints

1) Limit on the budget
2) Investment must be non-negative
3) There is a minimum return on the portfolio

* So the optimization problem is to choose a portfolio allocation that minimizes risk among all possible allocations that meet our requirements

eg:

$$\text{Max. } 0.5x + 0.4y + 0.35z$$

$$\text{st } x, y, z \geq 0$$

$$x + y + z \leq 12000$$

② Device Sizing

* Task of choosing the length of width of each device in electronic circuit

* Variable x represent the length ~~of~~ or width (device size) of the device

* Constraint

    * limits on the device sizes
    * limit on the total area of the circuit
    * timing that the circuit can operate reliably at a specified speed

* The optimization problem is to find the device size that satisfy the design requirements (timing, area, manufacturabilit

③ Data fitting:-

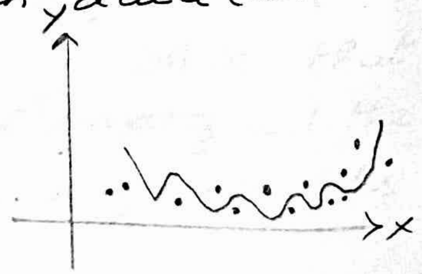
Model ~~filtering~~ fitting is the essence of machine learning

If one model does not fit the data correctly, the outcome it produces will not be accurate enough to be useful for decision making

Overfitting and Underfitting are the 2 important causes for the poor performance of ML algorithm

Overfitting:

When the model performs well on the training data but does not perform well on the evaluation data. It happens because its trained with so much data (including noise



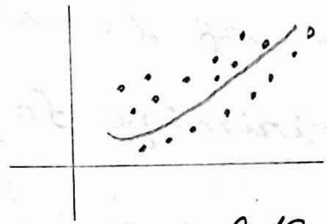
Underfitting:

* When the model is unable to capture the relationship between the input and output variables accurately. It generates error rate both on training set and unseen set

* Occurs because the data available to build a model is very less

Good Fit:-

* When the model makes the prediction with 0 error, it is said to have a good fit on the data



* This situation is achievable at a spot between overfitting and underfitting

* The optimization problem is to find the model parameter values that are consistent with the prior information and give the smallest misfit or prediction error with the observed data.

* Other areas of application

1) SCM, Finance, N/W design and operation

Advantages of Mathematical Optimization

* It is used as an aid to a human decision maker, system designer, or system operator who supervises the process, checks the result and modifies the problem when necessary

* The human decision maker also creates carries out any actions suggested by the optimization problem.

eg: Buying or selling assests to achieve the optimal portfolio

Solving Optimization problems:

Two important algorithms that solve convex optimization problems

1) Least squares
2) Linear Programming

Least square:

A least square problem is an optimization problem with no constraints (i.e $m=0$) and objective is a sum of square of terms of the form $a_i^T x - b_i$

$$\text{minimize } f_0(x) = \|Ax - b\|_2^2$$
$$= \sum_{i=1}^{k} (a_i^T - b_i)^2$$

Here $A \in R^{k \times n}$

rows of A., vector (with $k \geq n$) a_i^T are the

Definition:

It is a process of find the best fitting curve or line of best fit for a set of data points by reducing the sum of the square of the offset of the points from the curve

During the process of finding the relation between 2 variables, the trend of outcomes are estimated quantitatively. This is called regression analysis

Least Square Approximation

For data points $\{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$

* the matrix of linear function can be written as

$$A_1 = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$$

* The matrix of quadratic functions can be written as

$$A_2 = \begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & & \\ x_n^2 & x_n & 1 \end{pmatrix}$$

To calculate the error of linear function

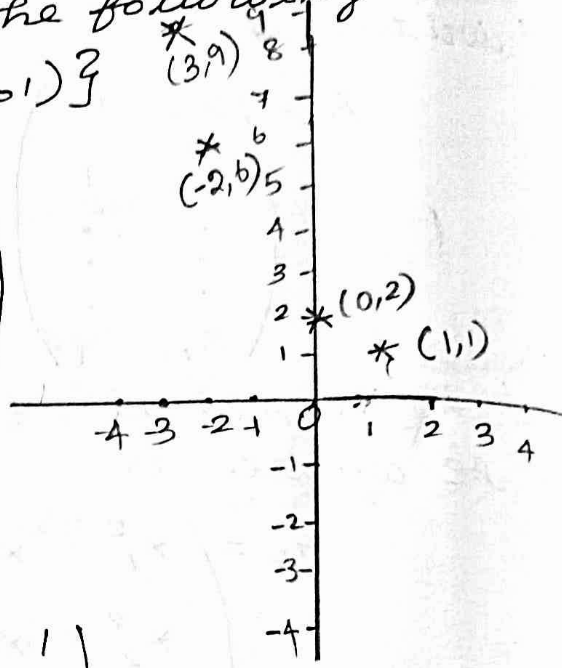
$$\boxed{E = \|A_1 x_0 - b\|^2, \; x_0 = (A_1^T A_1)^{-1} (A_1^T, b)}$$

Quadratic

$$E = \|A_2 x_0 - b\|^2, \; x_0 = (A_2^T A_2)^{-1} (A_2^T b)$$

Problems:

Use least square approximation to find the best fit with linear function. compute Error E for the following data
$\{(-3,9)\ (-2,6)\ (0,2)\ (1,1)\}$



Soln:

$$A_1 = \begin{pmatrix} -3 & 1 \\ -2 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \qquad y = \begin{pmatrix} 9 \\ 6 \\ 2 \\ 1 \end{pmatrix}$$

$$= A_1^T = \begin{vmatrix} -3 & -2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

$$= A_1^T A_1 = \begin{vmatrix} -3 & -2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} \begin{pmatrix} -3 & 1 \\ -2 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 9+4+1 & -3-2+1 \\ -3-2+1 & 1+1+1+1 \end{pmatrix} \qquad A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= A_1^T A_1 = \begin{pmatrix} 14 & -4 \\ -4 & 4 \end{pmatrix}$$

$$56 - 16 = 40$$

$$= \frac{1}{40} \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1/10 & 1/10 \\ 1/10 & 7/20 \end{bmatrix}$$

$$= (A_1^T A_1)^{-1} = \frac{1}{|A|} \text{adj } A$$

$$= \frac{1}{10} \begin{pmatrix} 1 & 1 \\ 1 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{7}{20} \end{pmatrix}$$

$$x_0 = (A_1^T A_1)^{-1} (A_1^T y)$$

$$= \begin{pmatrix} 1/10 & 1/10 \\ 1/10 & 7/20 \end{pmatrix} \begin{pmatrix} -3 & -2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ 6 \\ 2 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1/10 & 1/10 \\ 1/10 & 7/20 \end{pmatrix} \begin{pmatrix} -27-12+1 \\ 9+6+2+1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{7}{20} \end{pmatrix} \begin{pmatrix} -38 \\ 18 \end{pmatrix}$$

$$= \frac{-76+126}{20}$$

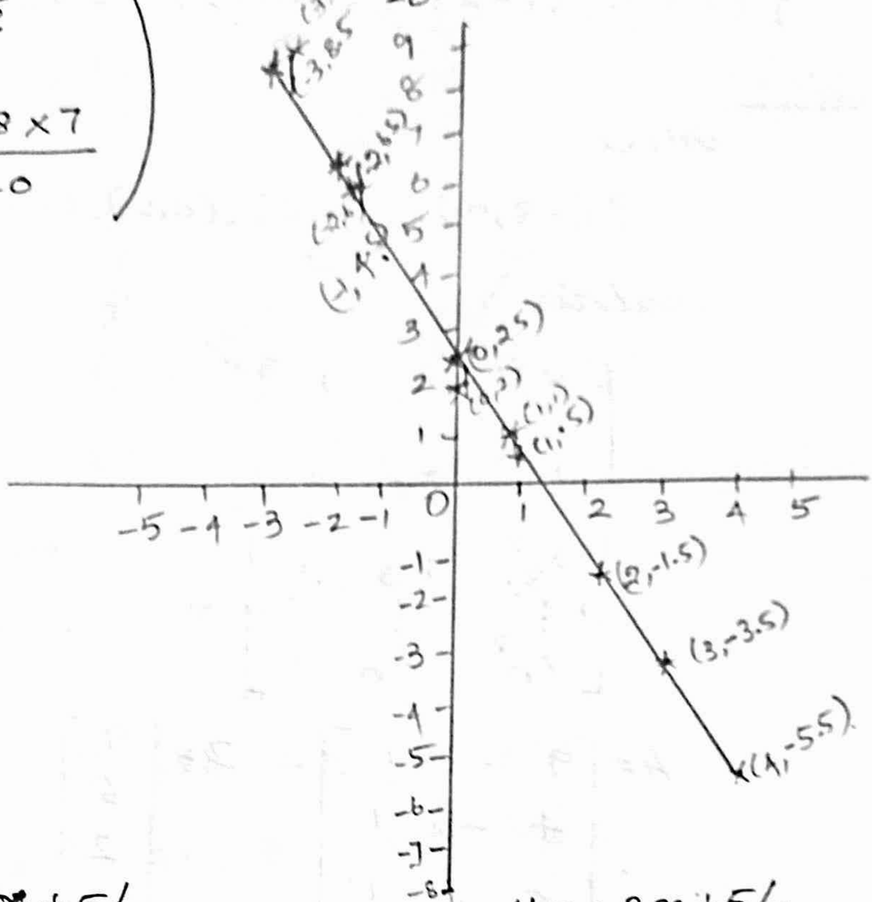
$$= \frac{50}{20}$$

Geometric interpretation.

$$= \begin{pmatrix} \frac{-38+18}{10} \\ \frac{-38+18\times 7}{10 \quad 20} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-20}{10} \\ \frac{50}{20} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-20}{10} \\ \frac{25}{10} \end{pmatrix}$$



$y = -2x + 5/2$

Best fit

ML model

$$y = -2\cancel{x} + 5/2$$

$$(A_1 x_0 - y) = \begin{pmatrix} -3 & 1 \\ -2 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 5/2 \end{pmatrix} \begin{pmatrix} 9 \\ 6 \\ 2 \\ 1 \end{pmatrix}$$

when $x = 1$ $y = .5$
$x = 2$ $y = -1.5$
$x = 3$ $y = -3.5$
$x = 4$ $y = -5.5$
$x = -1$ $y = 4.5$
$x = -2$ $y = 6.5$
$x = -3$ $y = 8.5$
$x = 0$ $y = 2.5$

$$= \begin{pmatrix} \frac{-1}{2} \\ 1/2 \\ 1/2 \\ -1/2 \end{pmatrix}$$

$$E = ||A_1 x_0 - y||^2$$

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

$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$$

$$E = 1$$

2) Use least square approximation find the best fit with quadratic function computes error E for the following data

$\{(-3,9),(-2,6),(0,2),(1,1)\}$

Soln:

Given

$\{(-3,9),(-2,6),(0,2),(1,1)\}$

Matrix

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \\ x_n^2 & x_n & 1 \end{bmatrix}$$

$$A=\begin{bmatrix} 9 & -3 & 1 \\ 4 & -2 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad D=\begin{bmatrix} 9 \\ 6 \\ 2 \\ 1 \end{bmatrix}$$

$$E=\|Ax_0-b\|^2$$

$$x_0=(A^TA)^{-1}A^Tb$$

$$A^TA=\begin{bmatrix} 9 & 4 & 0 & 1 \\ -3 & -2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 9 & -3 & 1 \\ 4 & -2 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 81+16+0+1 & -27+(-8)+0+1 & 9+4+0+1 \\ -27+(-8)+0+1 & 9+4+0+1 & -3+(-2)+0+1 \\ 9+4+0+1 & -3+(-2)+0+1 & 1+1+1+1 \end{bmatrix}$$

$$A^TA=\begin{bmatrix} 98 & -34 & 14 \\ -34 & 14 & -4 \\ 14 & -4 & 4 \end{bmatrix}$$

$$(\text{coeff } A)^T = \begin{vmatrix} 56-16 & -(-136+56) & 136-196 \\ -(-136+56) & +(392-196) & -(-392+476) \\ 136-196 & -(-392+476) & 1372-1156 \end{vmatrix}$$

$$= \begin{bmatrix} 40 & +80 & -60 \\ 80 & 196 & 84 \\ -60 & -84 & 216 \end{bmatrix}$$

$$\text{adj } A = \begin{bmatrix} 40 & 80 & -60 \\ 80 & 196 & -84 \\ -60 & -84 & 216 \end{bmatrix}$$

$$|A| = \begin{vmatrix} 98 & -34 & 14 \\ -34 & 14 & -4 \\ 14 & -4 & 4 \end{vmatrix}$$

$$\Rightarrow 98(56-16) + 34(-136+56) + 14(136-196)$$

$$\Rightarrow 98(40) + 34(-80) + 14(-60)$$

$$|A| \Rightarrow 3920 - 2720 - 840 = 360$$

$$A^T b \Rightarrow \begin{bmatrix} 9 & 4 & 0 & 1 \\ -3 & -2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 6 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 81+24+0+1 \\ -27-12+0+1 \\ 9+0+2+1 \end{bmatrix} = \begin{bmatrix} 106 \\ -38 \\ 18 \end{bmatrix}$$

$$x_0 = \frac{1}{360} \begin{bmatrix} 40 & 80 & -60 \\ 80 & 196 & -84 \\ -60 & -84 & 216 \end{bmatrix} \begin{bmatrix} 106 \\ -38 \\ 18 \end{bmatrix}$$

$$= \frac{1}{360} \begin{bmatrix} 4240 - 3040 - 1080 \\ 8480 - 7448 - 1512 \\ -6360 + 3192 + 234 \end{bmatrix}$$

$$= \frac{1}{360}\begin{bmatrix} 120 \\ -480 \\ 720 \end{bmatrix}$$

$$X_0 = \begin{bmatrix} 1/3 \\ -4/3 \\ 2 \end{bmatrix}$$

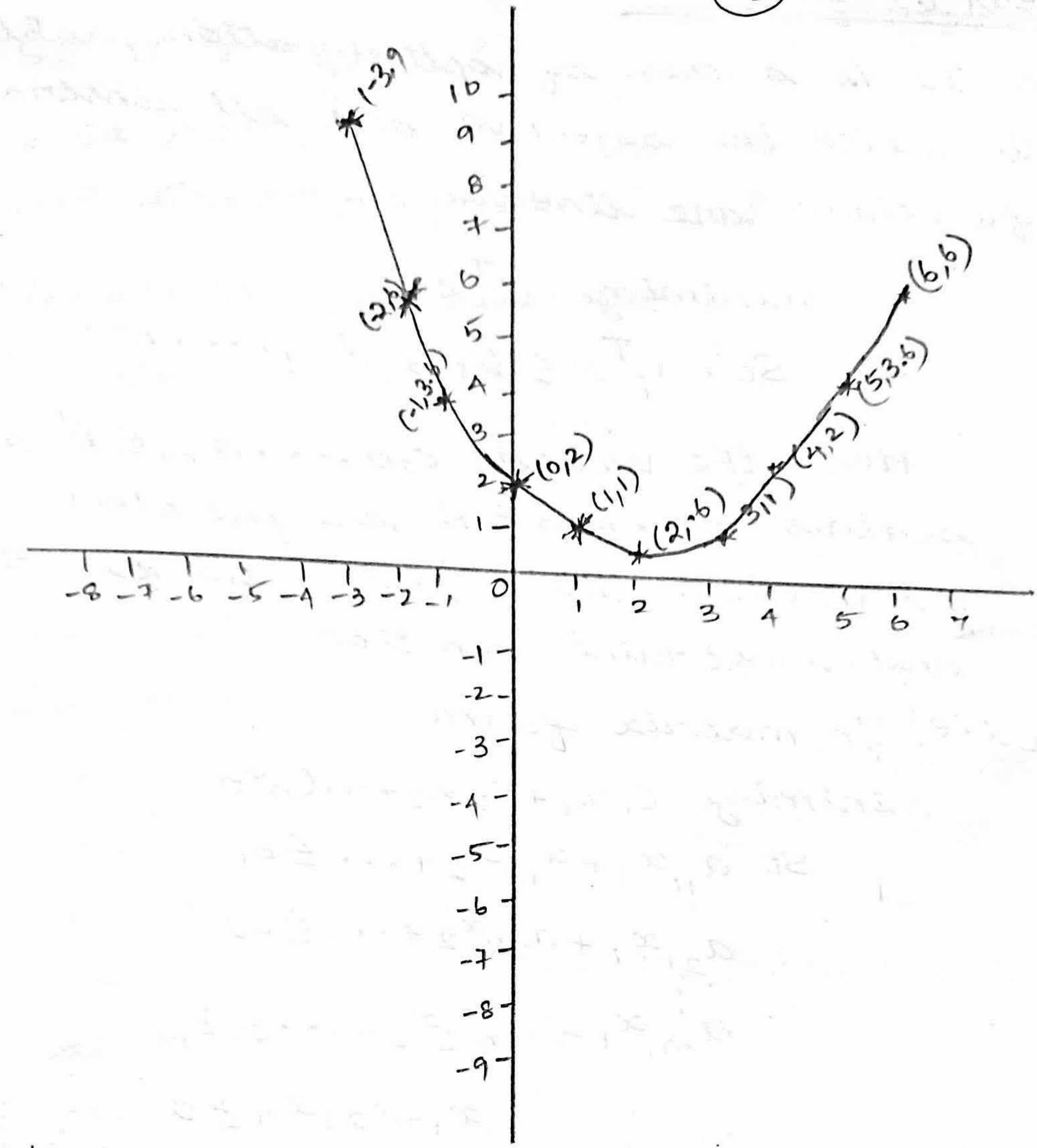
$$E = ||AX_0 - b||^2$$

$$AX_0 - b = \begin{bmatrix} 9 & -3 & 1 \\ 4 & -2 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1/3 \\ +4/3 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 2 \\ 1 \end{bmatrix}$$

$$AX_0 - b = \begin{bmatrix} 9 \\ 6 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 9 \\ 6 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$E = ||AX_0 - b||^2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$y = \frac{1}{3}x^2 - \frac{4}{3}x + 2 \Rightarrow$ **Best of fit of the equation**



No errors.

$$y = \frac{1}{3}x^2 - \frac{4}{3}x + 2$$

$x = -2 \quad y = \frac{1}{3}(4) - \frac{4}{3}(-2) + 2 = 6$

$x = -1 \quad y = \frac{1}{3}(+1) - \frac{4}{3}(-1) + 2 = 3.6$

$x = 0 \quad y = 2$

$x = 1 \quad y = \frac{1}{3}(1) - \frac{4}{3}(1) + 2 = 1$

$x = 2 \quad y = \frac{1}{3}(4) - \frac{4}{3}(2) + 2 = 0.6$

$x = 3 \quad y = \frac{1}{\cancel{3}}(\overset{1}{\cancel{3}}) - \frac{4}{3}(3) + 2 = 1$

$x = 4 \quad y = \frac{1}{3}(16) - \frac{4}{4}(4) + 2 = 2$

$x = 5$

$y = \frac{1}{3}(25) - \frac{4}{3}(5) + 2$

$= 3.6$

$x = 6$

$y = \frac{1}{3}(36) - \frac{4}{3}(6) + 2 = 6$

$x = -3.$

$y = \frac{1}{3}(-3)^2 - \frac{4}{3}(-3+2)$

$= 9$

Linear Programming

It is a class of optimization problem in which the objective and all constraint functions are linear

$$\text{minimize } c^T x$$
$$\text{st. } a_i^T x \leq b_i, \quad i = 1 \dots m$$

Here the vectors $c, a, \dots a_m \in R^n$ and scalars $b_1 \dots b_m \in R$ are problem parameters that specify the objective and constraint function

(i.e) In matrix form

$$\boxed{\begin{aligned} \text{minimize } & c_1 x_1 + c_2 x_2 + \dots c_n x_n \\ \text{st } & a_{11} x_1 + a_{12} x_2 + \dots \leq b_1 \\ & a_{21} x_1 + a_{22} x_2 + \dots \leq b_2 \\ & \vdots \\ & a_{m1} x_1 + a_{m2} x_2 + \dots \leq b_m \\ & x_1, x_2, x_n \geq 0 \end{aligned}}$$

eg:

$$\boxed{\begin{aligned} \min \quad & z - 3x_1 + 4x_2 \\ \text{st.} \quad & x_1 + x_2 \geq 5 \\ & 3x_1 + x_2 \geq 8 \\ & x_1, x_2 \geq 0 \end{aligned}}$$

$$c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}^T \quad (\text{or}) \quad c = [c_1 c_2 \dots c_n], \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$a = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$\therefore$ min $c^T x$

St. $a_i^T x \le b_i$, $i = 1 \dots m$

Simplex Method is an effective method to solve linear programming problem

Ex: Chebyshev Approximation problem

minimize $\max_{i=1 \dots K} [a^T x - b_i]$

when $x \in R^n$ is the variable

$a_1, \dots a_K \in R^n$, $b_1, \dots b_K \in R^n$ are parameters that specify the problem instance

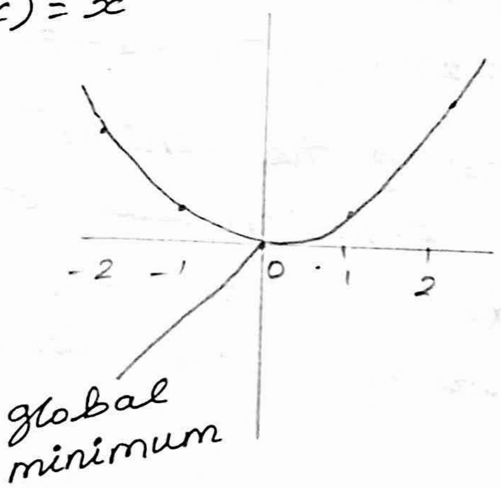
* The objective is to maintain and measure the maximum of the absolute values

* Here the objective function is not differentiable

<u>Convex Optimization</u>

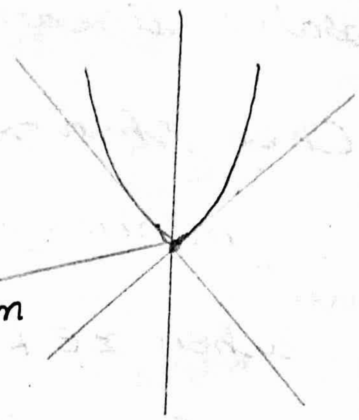
Lets take the single variable function $y = x^2$. This has only one local minimum at $x = 0$ and hence has one global minimum at that point.

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



x	y
0	0
1	1
2	4
-1	1
-2	4

Lets take multi variable function $z = x^2 + y^2$. It also has only one local minimum at (0,0) and has one global minimum at that point.



global minimum

It is easy to find the location of these and more complex multivariable functions, that has only one global minimum

* When these multivariate function have only one local minimum and hence that becomes the global minimum, the graph of these functions share a geometrical property called convexity Such functions are called convex function.

* But in Pratice, we need to find the minimum value of the functions subject to constraints expressed by
eg: $2x + 3y = 7$, $5x - 8y <= 8$

* Finding the minimum value of the convex function subject to constraint like above is the topic of a field of study called convex optimization

* Gradient Descent Algorithm is the famous algorithm that is used to solve all convex optimization problems very fast

* But in Engineering these are lot of functions we need to minimize that are not convex.

Non - convex functions:-

eg: $y = x^4 + 2x^3 + 5x^2 + 9x + 1$

$z = x^4 + y^4 + x^3 + y^3$

* A non convex function will result in different local minima. depending on the realization of the input data.

* But a convex function is more like likely to result in the same or similar minima.

Therefore it is necessary to convert a non-convex problem to a convex problem inorder to get one global minimum

Need to transform from non-convex to convex

* As minimizing a non-convex function subject to constraint is very difficult

* There is no universal method that works for every function

* There are only few algorithm that work well for a small group of non-convex function

This topic is called non-convex optimization

Stnd form

A convex optimization problem is one of the form

minimize $f_0(x)$
s.t. $f_i(x) \leq b$, $i = 1 \ldots m$
where the function $f_0 \ldots f_n : R^n \to R$ are convex (ie) satisfy

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$
$\forall x, y \in R^n$ and $\forall \alpha, \beta$ with $\alpha + \beta = 1$,
$\alpha \geq 0$, $\beta \geq 0$

* Least square problem and Linear Programming are both special causes of the general convex optimization problem

* Interior Point methods work very well. with good accuracy

Interior Point Method

* Very reliable
* It can easily solve hundreds of variables and thousands of constraints on a desktop computer in at most a few ten of seconds

Advantages of convex optimization:

* If we can formulate a problem as a convex optimization then we can solve it very efficiently

* If we formulate a practical problem as a convex optimization problem then you have solved the original problem.

Challenges:

* Recognizing a convex function is difficult

* There are many tricks needed for transforming a convex problem. Hence challenging

* Once the skill of recognizing or formulating convex problem optimization is developed. Then many problems can be solved by convex optimization

Non-linear optimization

* Otherwise called as non-linear programming is ther term used to describe an optimization problem when the objective or constraint functions are not linear but are not convex

* Unfortunately, there are not no effective methods for solving the general non linear programming problem

* Even problems with few ten. Variables can be extremely challenging.

Local Optimization

* In local optimization, the compromise is to give up selecting the optimal 'x' that minimizes the objective over all feasible points

* This selects and seeks a point that is only local optimal (i.e) It minimizes the OF among the feasible point that are near to it

* It does not guarenteed to have a lower objective value than all other feasible points

Advantages:

* It is fast
* Can handle large scale problems
* It is widely used in applications where there is value in finding good point if not the very best

Disadvantages:

* It does not find the true globally optimal soln.

* It requires an initial guess for the optimization variable

* This initial guess is critical and can largely affect the objective value of the local solution obtained

* More sensitive to algorithm parameter values which may need to be adjusted for a particular problem or family of problems

* It involves good choice of algorithms, good initial guess; adjusting algorithm parameter

Global Optimization:

* Here the ~~two~~ true global solution of the optimization problem is found
* Even problems with a few ten of variables can take a very long time to solve
* It can be used for problems with a small number of variables where computing time is not critical and the value of finding the true global soln is very high
* It will find the absolute worst values of the parameters and if associated performance is acceptable, can certify the system as safe

Disadvantages:

* computing cost is high. but worth where the values of certifying the performance is high or the cost of being wrong about the reliability or safety is high

Convexity:

* A convex optimization problem is an optimization problem where you want to find a point that maximizes/minimizes the OF through iterative computations involving convex functions.

* The OF is subjected to equality and inequality constraints

* Convex optimization can be used to optimizes an algorithm that will increase the speed at which the algorithm converges to the solution

* To solve convex optimization problems ML techniques such as GD are used.

* Plays an important role in convex optimization

* It ensures that convex optimization problem are smooth and have well defined derivatives to enable the use of GD

* For convexity convex sets are very important

* A convex set is a set that contains all points on or inside its boundary and contains all convex combination of points in its interior.

(25) * A convex set is defined as a set of all convex functions

* A convex optimization problem is thus to find the global maximum or minimum of convex functions.

Why convex optimization such a big deal in ML?

* Many methods in ML are based on finding parameters that minimize some OF

* Optimization is when we search for variable that attain a global maximum or minimum of some functions

* Convex optimization is a subset of optimization where the functions you work with all convex which means 'bowl shaped'. This makes the search for maxima and minima easier since you just walk on the surface of the bowl in which the direction with the greatest slope to get there.

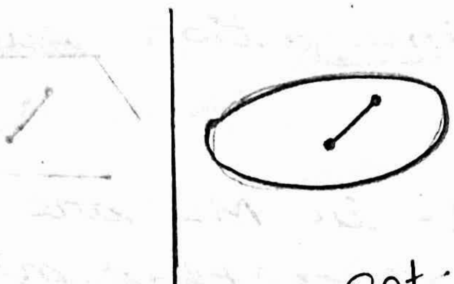
* That direction happens to be the gradient or derivative of the function describing the bowl

* Convex problems are guarenteed to have any minimum (global minimum)

* Concave problems are guarenteed to have any maximum (global maximum)

Advantages of convexity

*A convex feasible region makes it easier to ensure that you do not generate infeasible solutions while searching for the optimum

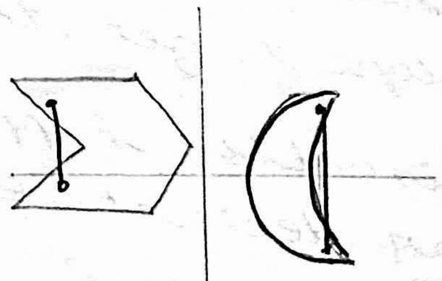


convex Set.

* With convex objective of convex feasible region ther can be only one optimal solution which is globally optimal

* Convex problems can be solved efficiently up to a very large size

*A non-convex optimization problem is any problem where the objective or any of the constraints are non convex



Not a convex set.

*These problems have multiple feasible region of multiple local points within the region

(27) Convex function Geometrically (Graph - Bowl shaped)

A function is convex if a line segment drawn from any point $(x, f(x)$ to another point $(x, f(y))$ called the chord from x to y lies on or above the graph of f.

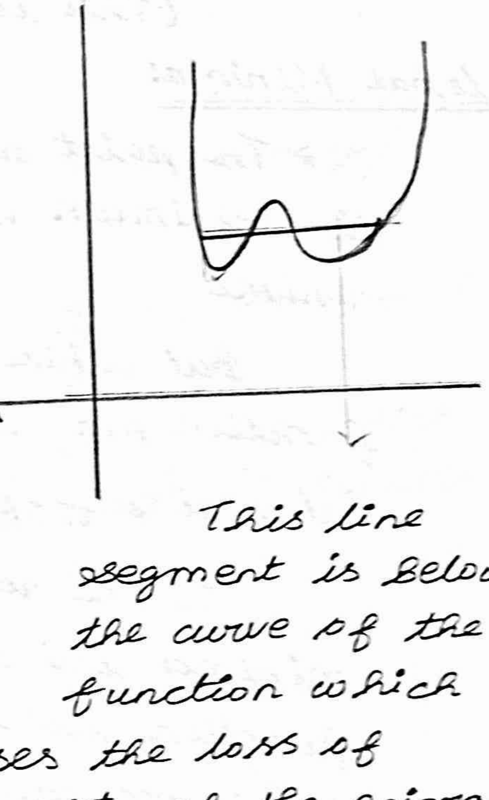
Every linear function whose graph is a straight is both convex and concave.

It is a epigraph in a convex set

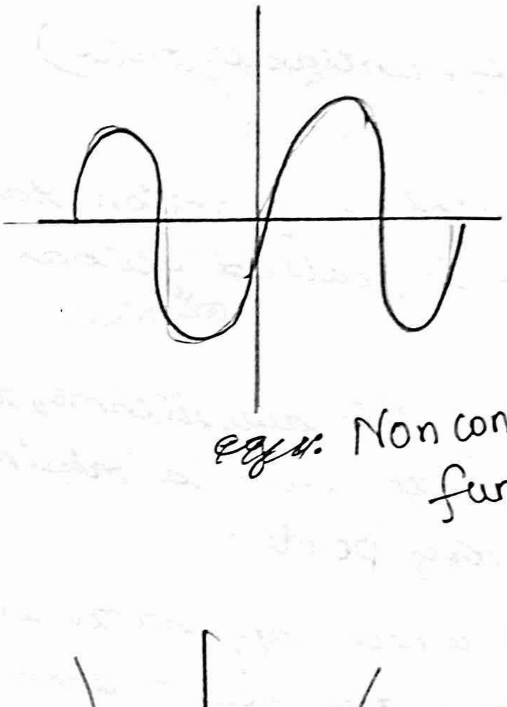
Non convex function:

A non-convex function curves up and down, if it is neither convex nor concave

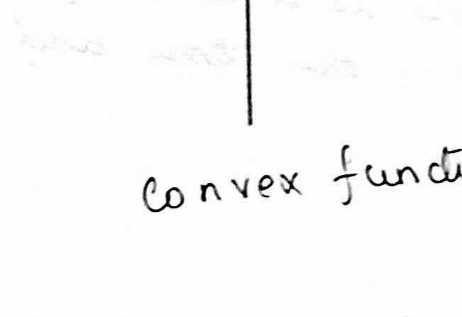
eg 2 Non convex function



eg 1. Non convex function



This line segment is Below the curve of the function which causes the loss of convexity of the epigraph



Convex function

How will you find whether a function is convex without plotting a graph

Test for convexity: (Mathematically)

* If the second order derivative of that function is greater than or equal to 0

$$f''(x) \geq 0$$

eg: $y = e^x$, $y = x^2$

(Both are differentiable twice)

* If $-f(x)$ is a convex function then the function is a concave function

$$-f''(x) \geq 0$$

eg:

$$y = -e^x$$

(This is differentiable twice)

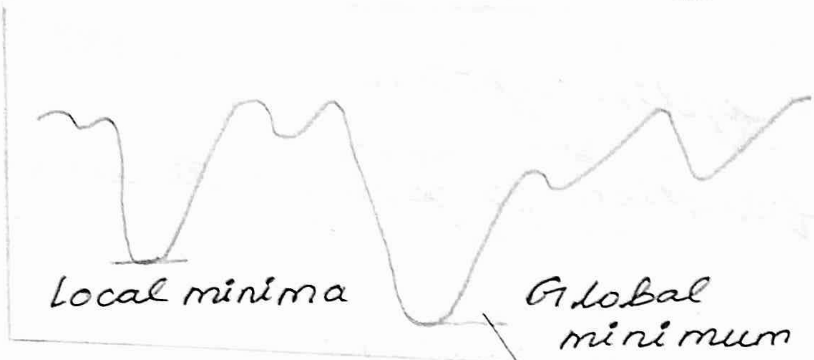
Global Minima:

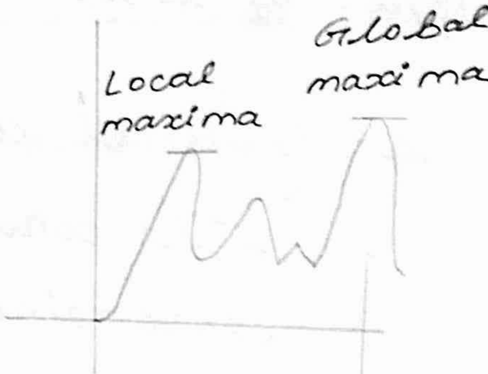
* The point at which a function takes the minimum value is called global minima

But while using GD algorithms, the function may appear to have a minimum value at diff. points.

Those points which appear to be minima but are not the point where the function actually takes the minimum value are called local minima

Gradient descent is able to find local minima most of the time and not global minima





$f''(x) > 0$

A function can have more than one local min and local max value

A function can have only one absolute (global) max and min value