DATA SCIENCE FOR ENGINEERS

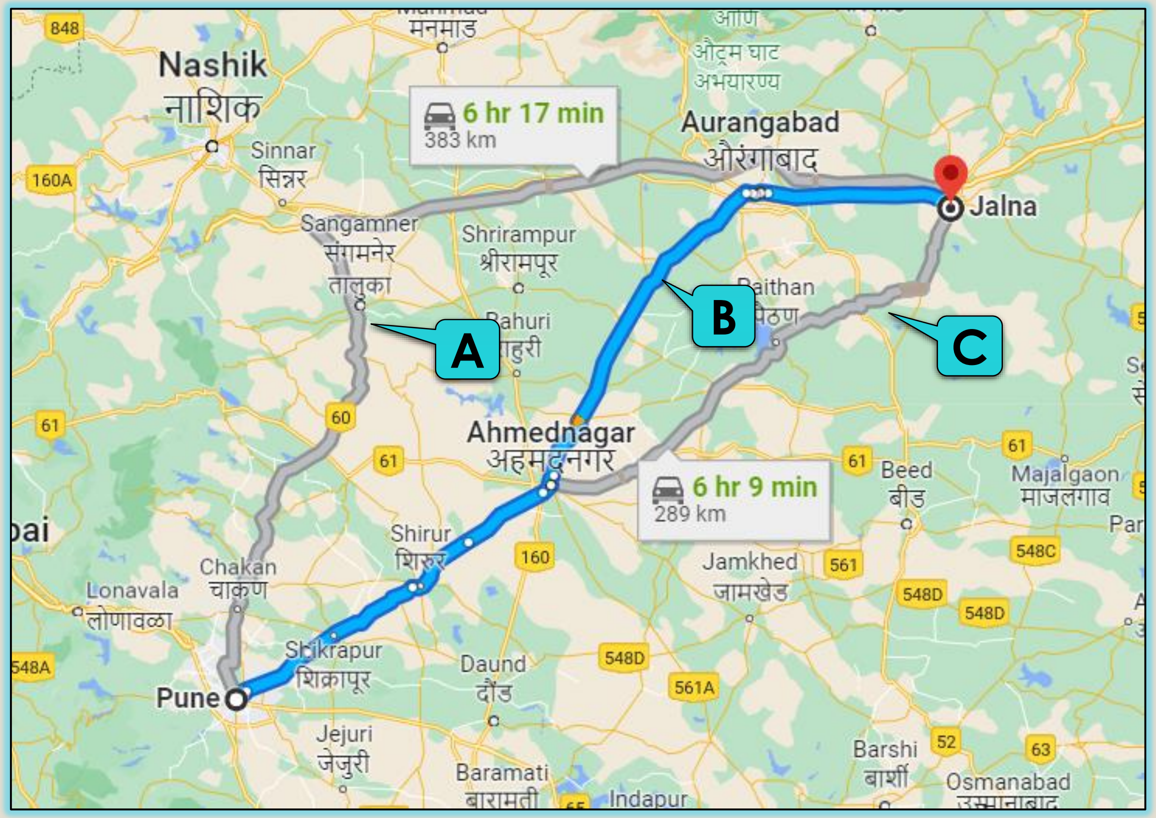
Week 4

Session Co-Ordinator : Abhijit Bhakte

Three Pillars of Data Science

Optimization

- A optimization problem consists of maximizing or minimizing a real function by systematically choosing input value from within as allowed set and computing the value of the function (Wikipedia)



Route A:
D = 383 Km
T = 6hr 17min

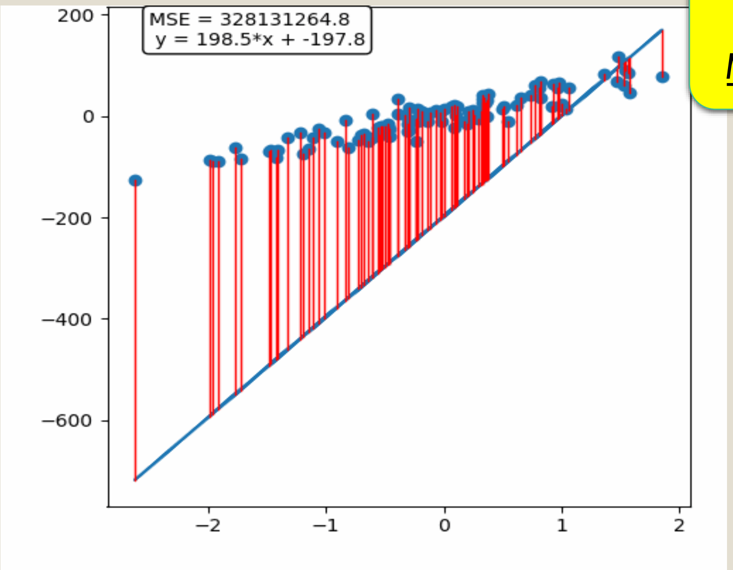
Route B: (optimal time path)
D = 294 Km
T = 6hr 3min

Route C: (optimal distance path)

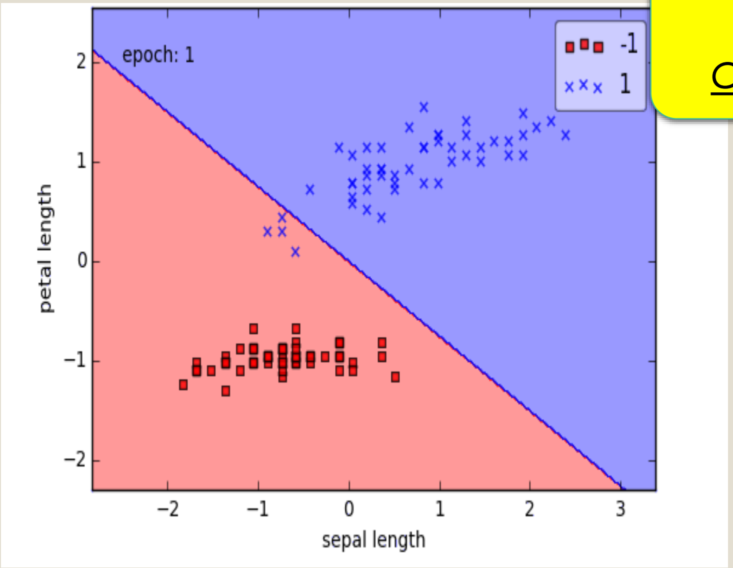
D = 389 Km
T = 6hr 9min

Optimization

➢ Optimization is the use of specific method to determine the ***'best'*** solution to the problem



Best line to
represent data



Best line to
classify data

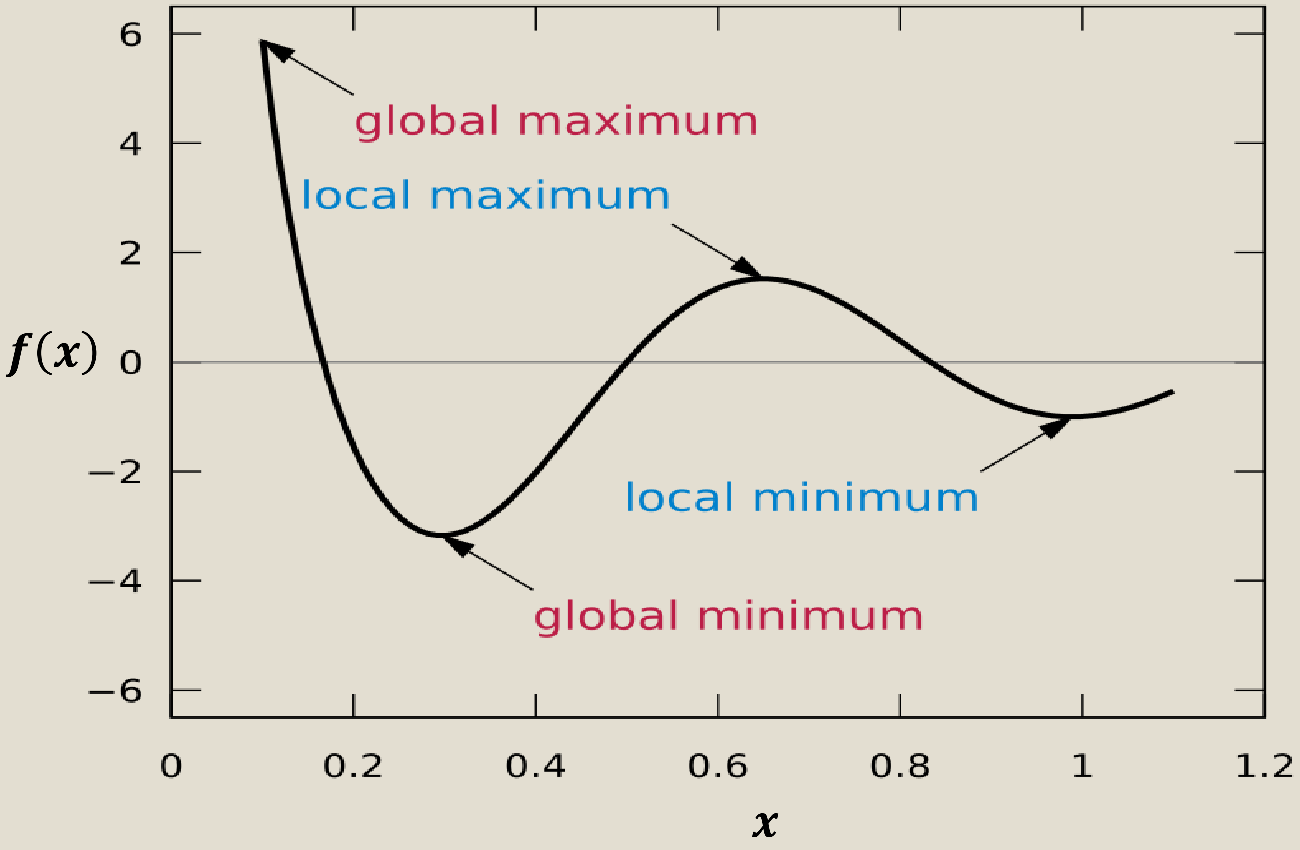
Univariate Optimization problem

- Objective function
- Decision variable
- Constraints

$$\min_{x} f(x)$$
$$x \in R$$

Decision variable

Objective function



Univariate optimization conditions

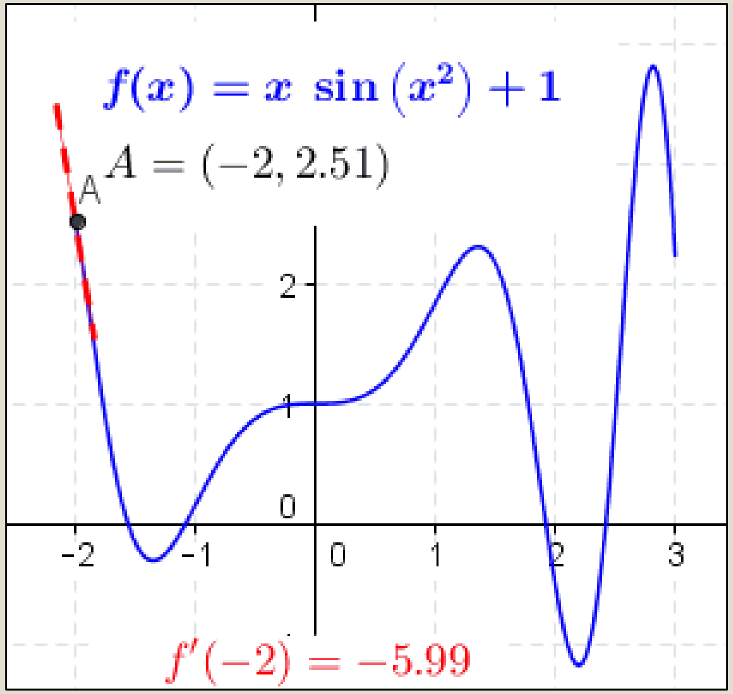
$$\min_{x} f(x)$$
$$x \in R$$

➢ Necessary condition for x to be minimizer

$$f'(x) = 0 \longrightarrow \frac{dy}{dx} = \frac{y_2 - y_1}{x_2 - x_1} = 0$$

➢ Sufficient condition

$$f''(x) > 0 \longrightarrow \frac{d^2y}{dx^2}$$



R Studio

Que: $f(x) = \frac{x^4}{4} - \frac{x^3}{3} - x^2$

a) Find the stationary points for the following points

b) Find the values of x at which minima exist and its value

- The first order necessary condition is $f'(x) = 0$

$$f'(x) = x^3 - x^2 - 2x = 0$$

➢ Solving above equation gives stationary points are: $\underline{\boldsymbol{x = 0,\ -1,\ +2}}$

- The second order necessary condition for minima is $f''(x) > 0$

$$f''(x) = 3x^2 - 2x - 2$$

- Putting stationary values in above equation

$f''(x = 0) = 3(0)^2 - 2(0) - 2 = -2$ → $\mathrm{f}(x = 0) = 0$

$\underline{\boldsymbol{f''(x = -1) = 3(-1)^2 - 2(-1) - 2 = 3}}$ → $\underline{\boldsymbol{f(x = -1) = -0.41667}}$

$\underline{\boldsymbol{f''(x = +2) = 3(2)^2 - 2(2) - 2 = 6}}$ → $\underline{\boldsymbol{f(x = +2) = -2.667}}$

➢ The minima exist at x = +2 , with the value = - 2.667

Q) In optimization, what does the term "local minimum" refer to?

A) The lowest point in the entire solution space
B) The lowest point within a specific region
C) The highest point in the solution space
D) The highest point within a specific region

Q) What is the minimum point of the function $f(x) = x^2 - 4x + 5$?

A) X = 2, f(x) = 1
B) X =1, f(x) = 2
C) X = 2, f(x) =-1
D) X = -1, f(x) = 2

Explanation:

To find the minimum point, we take the derivative $f'(x)$ and set it equal to 0.
Step 1) $f'^{(x)} = 2x - 4 = 0$ → x = 2

Step 2) put x = 2 in equation $x^2 - 4x + 5$
$f(2) = 2^2 - 4 * 2 + 5 = 1$

Que: Off all rectangles of area 100, which has the smallest perimeter ?

- Let l be the length and b be breadth of the rectangle

$$Area = l \times b, \qquad Perimeter = 2l + 2b$$

- Here, the optimization function is perimeter and decision variable is length

$$f(l) = 2l + 2b = 2l + 2\left(\frac{A}{l}\right) = 2l + \frac{200}{l}$$

- Applying first order necessary condition $f'(x) = 0$

$$f'(l) = 2 - \frac{200}{l^2} = 0$$

- Solving above equation gives length = +10,-10
- As we know length is cannot be negative hence the length of the rectangle is 10.
- Calculating the breadth=10, the perimeter of the rectangle is ***P = 2(10)+2(10)= 40***

Que: A manufacturer determines that the daily avg of producing q units is $C(q) = 0.0001q^2 - 0.08q + 65 + (5000/q)$ determine the number of units produce per day which minimize the avg cost ?

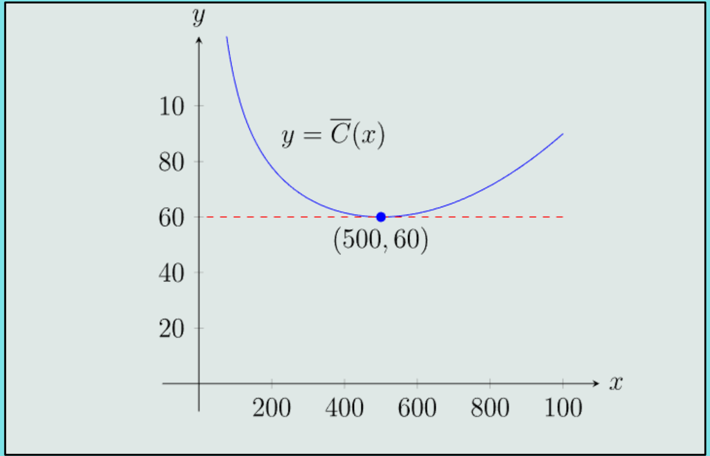
- Here, the optimization function is C(q) and q is the decision variable

$$C(q) = 0.0001q^2 - 0.08q + 65 + (\frac{5000}{q})$$

- Applying first order necessary condition $C'(q) = 0$

$$C'(q) = 0.0002q - 0.08 - (\frac{5000}{q^2})$$

- Solving above equation, the critical value **q= 500**



- Checking the second order necessity condition $C''(q) > 0$

$$C''(q) = 0.0002 + (\frac{10000}{q^3})$$

- Putting critical value into the equation give positive value $C''(q) > 0$. Hence the manufacturing cost can be minimize by producing **q = 500 units**

Q) A rectangle has a fixed perimeter of 24 units. What dimensions would maximize its area?

A) Length = 6 units, Width = 6 units
B) Length = 8 units, Width = 4 units
C) Length = 12 units, Width = 0 units
D) Length = 10 units, Width = 2 units

Explanation:

Perimeter = 2*length + 2* breadth =24
Step 1) 2*L+2*B=24 → L+ B=12 → L= 12-B

Step 2) forming optimization function f(x)=A
A = L*B =(12-B)*B = $12B - B^2$

Step 3) f'(x)= 0 → f'(x)=12-2B=0 → B= 6 , L= 6

Q) Consider the function $f(x) = x^3 - 6x^2 + 9x + 2$. What are the critical points?

A) X = 1, x = 5
B) X = 0, x = 6
C) X = - 3, x = 1
D) X = 1, x = 3

Explanation:

To find the critical point, we take the derivative$f'(x)$ and set it equal to 0.
Step 1) $f'^{(x)} = 3x^2 - 12x + 9 = 0$→

Step 2) Solving above equation we get put

$$x = 1, 3$$

Q) A company wants to maximize its profit $\mathrm{P} = -2x^2 + 40x + 100$, where x is the quality of products sold. What is optimal quantity of product to sell?

A) 5
B) 10
C) 15
D) 20

Explanation:

To find the maximum profit, we take the derivative $\mathrm{P}'(x)$ and set it equal to 0.

Step 1) $P'(x) = -4x + 40 \; = 0 \rightarrow \mathrm{x} = 10$

Q) A car travels along a straight road. Its position is given by $\mathrm{s(t)} = 2t^2 + 3t + 5$
What is the car's velocity at t=2 seconds

A) 19 m/s
B) 17 m/s
C) 13 m/s
D) 11 m/s

Explanation:

The velocity v(t) of the car is the derivative of the position function s(t) with respect to *t*. Taking the derivative, we get $\mathrm{v(t)} = 4t + 3$.

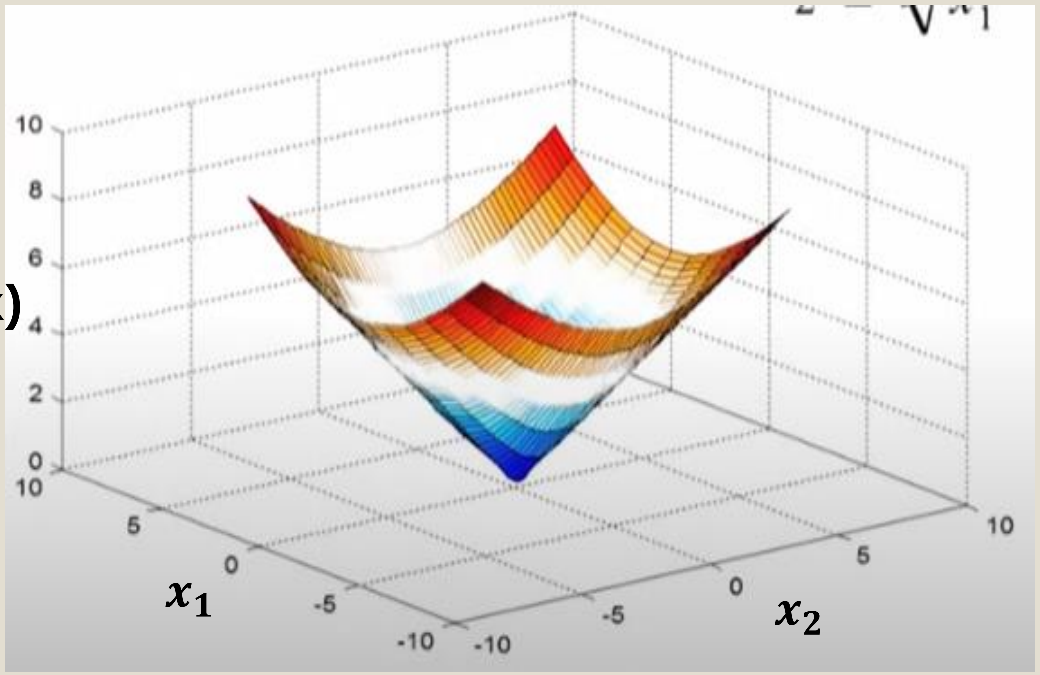
Plugging in *t*=2, we get $\mathrm{v(2)} = 4 * 2 + 3 = 11\; m/s$

Multivariate optimization

$$z = \mathrm{f}(x_1, x_2, .., x_n)$$

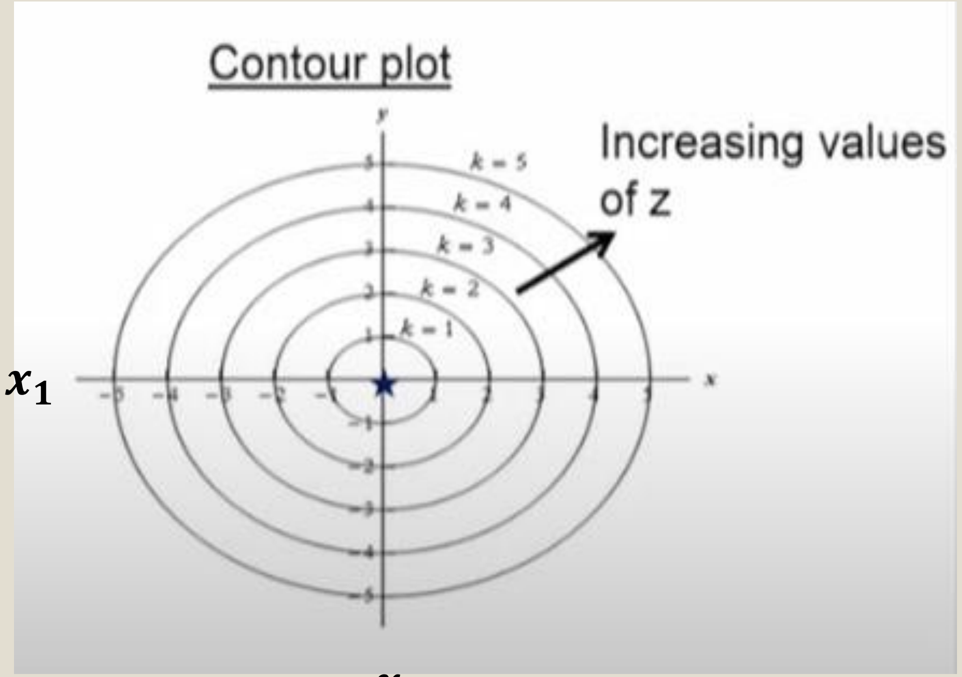
- Let $z = \sqrt{{x_1}^2 + {x_2}^2}$

f(x)



x_1 x_2

x_1



x_2

Q) The minima/maxima of f(x) exist when

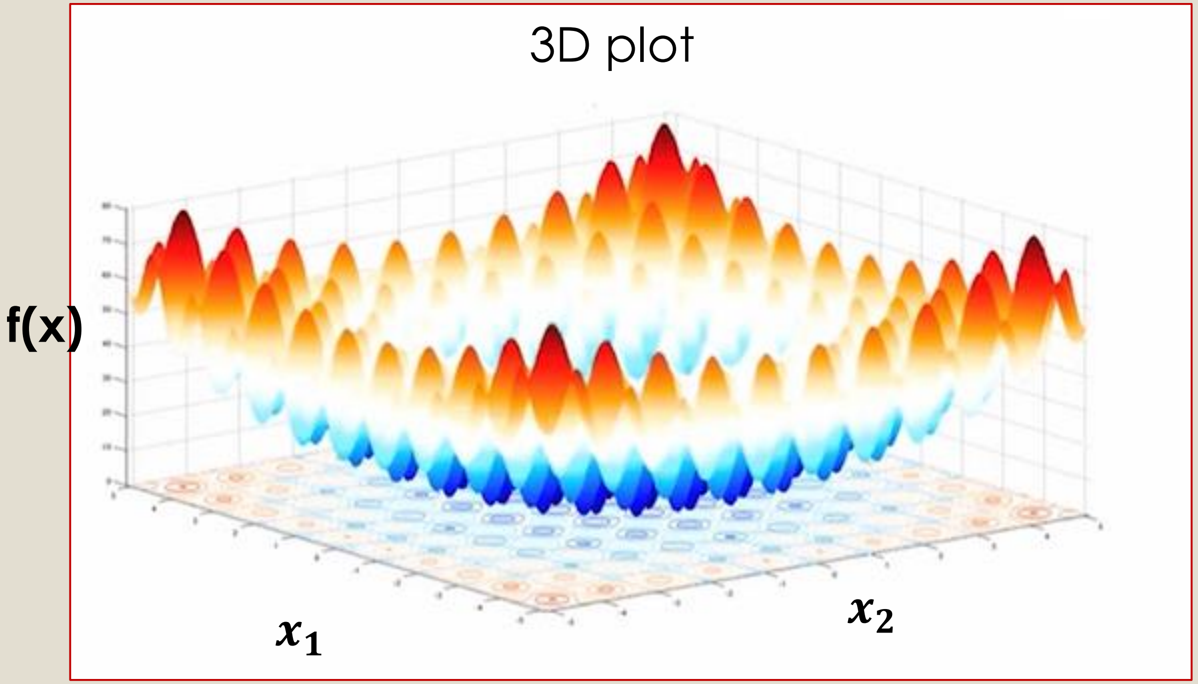
a) $f'(x) > 0$
b) $f'(x) = 0$
c) $f'(x) < 0$

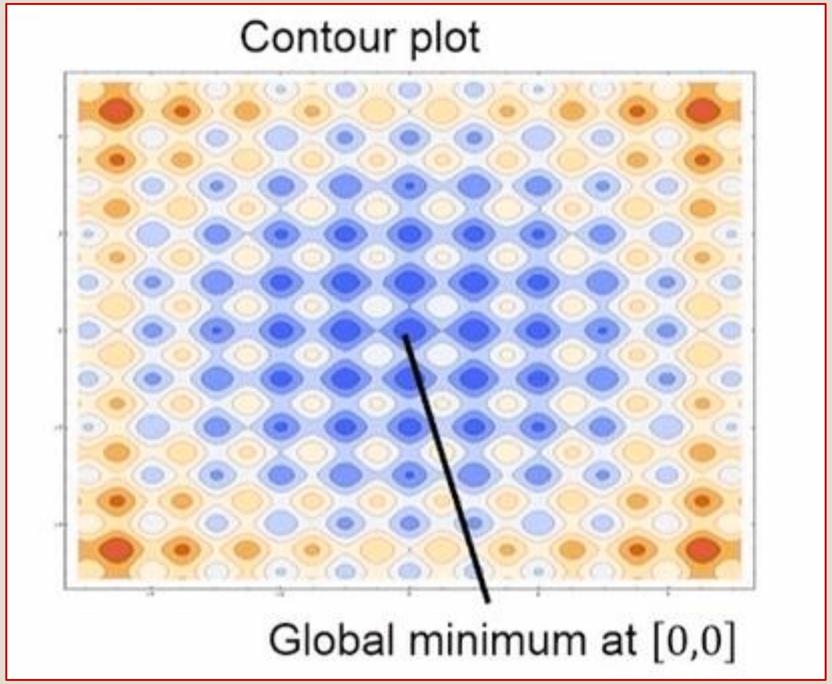
Q) The maxima of f(x) exist when

a) $f''(x) > 0$
b) $f''(x) = 0$
c) $f''(x) < 0$

Multivariate local and global minimum

$$f(x_1, x_2) = 20 + \sum_{i=1}^{2}[x_1{}^2 - 10\cos(2\pi x_i)]$$





Multivariate optimization condition

$$z = \mathrm{f}(x_1, x_2, .., x_n)$$

- Necessary condition for x to be minimizer : $\boldsymbol{\nabla f(x^*) = 0}$

$$\nabla f(x^*) = Gradient = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ ... \\ ... \\ \partial f/\partial x_n \end{bmatrix}$$

- Sufficient condition : $\boldsymbol{\nabla^2 f(x^*)}$ **has to positive definite (matrix with positive Eigen values)**

$$\nabla^2 f(x^*) = Hessian = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & ... & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & ... & \partial^2 f/\partial x_2 \partial x_n \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & ... & \partial^2 f/\partial x_n^2 \end{bmatrix}$$

Que: find the 1st and 2nd order necessary conditions for the function and tell weather minima exist or not ?

- The function is $f(x_1, x_2) = x_1 + 2x_2 + 4{x_1}^2 - x_1 x_2 + 2{x_2}^2$
- Applying first order necessary condition $\nabla f = 0$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 + 8x_1 - x_2 \\ 2 - x_1 + 4x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix} = \begin{bmatrix} \mathbf{-0.19} \\ \mathbf{-0.54} \end{bmatrix}$$

- Checking the second order necessity condition i.e. $\boldsymbol{\nabla^2 f(x^*)}$ ***is positive definite***

$$\boldsymbol{\nabla^2 f} = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial {x_2}^2} \end{bmatrix} = \begin{bmatrix} 8 & -1 \\ -1 & 4 \end{bmatrix}$$

- The eigen values of the $\boldsymbol{\nabla^2}$f matrix ***are 3.76 and 8.23 > 0*** . Hence the matrix is positive definite and ***minima exist at the critical points*** .

R Studio

Q) In multivariate optimization, what is the critical point of a function?
A) A point where the function is undefined
B) A point where the function has a local minimum
C) A point where the gradient is zero
D) A point where the function is linear

Explanation:

A critical point in multivariate optimization is a point where the gradient of the function is zero. It can be a local minimum, maximum, or a saddle point.

Q) In multivariate optimization, what is the role of the Hessian matrix?
A) It represents the gradient of the function
B) It helps in finding the critical points
C) It provides information about the function's curvature
D) It is used to evaluate partial derivatives

Explanation:

The Hessian matrix contains second partial derivatives and provides information about the curvature of the function at a critical point, helping to determine its nature.

Q) Calculate gradients for $f(x,y) = 4x^2 - 3xy + 6y^2$ at point (0,1).

A) [-3 ; 12]
B) [8 ; -3]
C) [8 ; 12]
D) [12 ; 8]

Explanation:

Step1) calculate derivatives wrt x and y

$\frac{df}{dx} = 8x - 3y;\quad \frac{df}{dy} = -3x + 12y;$

Step 2) put x= 0 and y=1in above equation gives gradients

$\boldsymbol{\frac{df}{dx} = 8(0) - 3(1) = -3;\quad \frac{df}{dy} = -3(0) + 12(1) = 12;}$

Q) Calculate hessian matrix for $f(x,y) = 4x^2 - 3xy + 6y^2$ at point (0,1).

A) $\begin{bmatrix} -8 & 3 \\ 3 & 12 \end{bmatrix}$ B) $\begin{bmatrix} 8 & -3 \\ -3 & 12 \end{bmatrix}$

C) $\begin{bmatrix} 12 & 2 \\ 2 & 8 \end{bmatrix}$ D) $\begin{bmatrix} -12 & -2 \\ -2 & 8 \end{bmatrix}$

Explanation:

Step1) calculate double derivatives wrt x and y

$\frac{d^2f}{dx^2} = 8;\quad \frac{d^2f}{dx\,dy} = -3;\quad \frac{d^2f}{dy^2} = 12$

$$\boldsymbol{\nabla^2 f} = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial {x_2}^2} \end{bmatrix} = \begin{bmatrix} 8 & -3 \\ -3 & 12 \end{bmatrix}$$

Q) A function $f(x, y) = x^2 + 4y^2$ has a critical point at (2,−1). What is the value of the function at this critical point?

A) 8
B) 14
C) 18
D) 20

Explanation:

Plugging x = 2 and y = -1 into function $f(x, y) = x^2 + 4y^2$, we get $f(2, -1) = 2^2 + 4(-1)^2 = 8$

Q) A quadratic function $f(x, y) = 3x^2 - 2xy + 2y^2$, What can be said about the critical point?

A) It's a local maximum
B) It's a local minimum
C) It's a saddle point
D) It's a global maximum

Explanation:

Step1) calculate derivatives wrt x and y

$\frac{d^2 f}{dx^2} = 6; \quad \frac{d^2 f}{dx\,dy} = -2; \quad \frac{d^2 f}{dy^2} = 4$

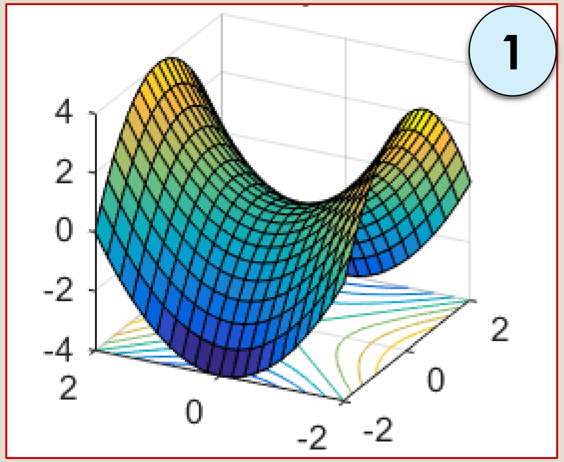
Step 2) calculate eigenvalues of Hessian matrix

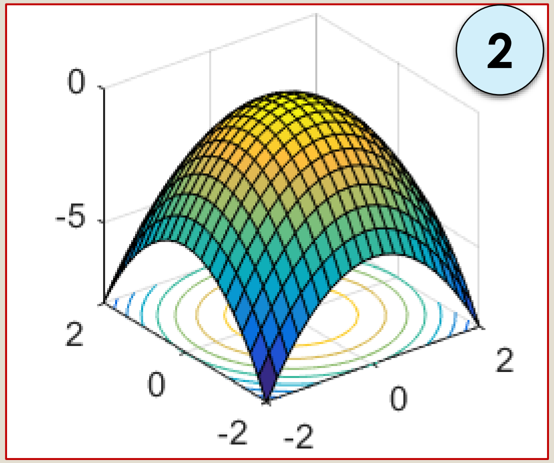
$\mathrm{H} = \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}; \rightarrow (\mathrm{H} - \lambda \mathrm{I}) = 0 \rightarrow \lambda^2 - 10\lambda + 20 = 0$

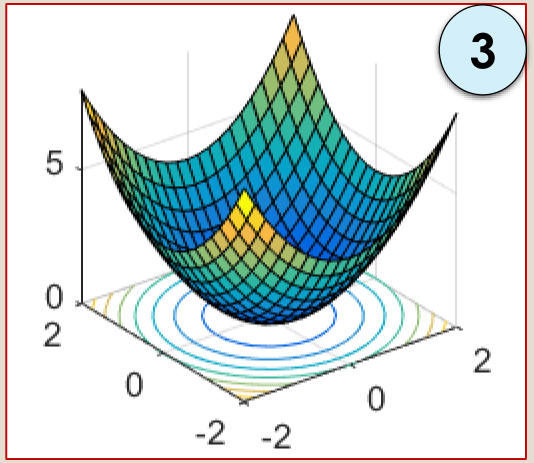
λ = 5 $\pm \sqrt{5}$

As eigenvalues are positive hence matrix is positive definite and represent existence of local minima

Q) Choose the correct option for given figure



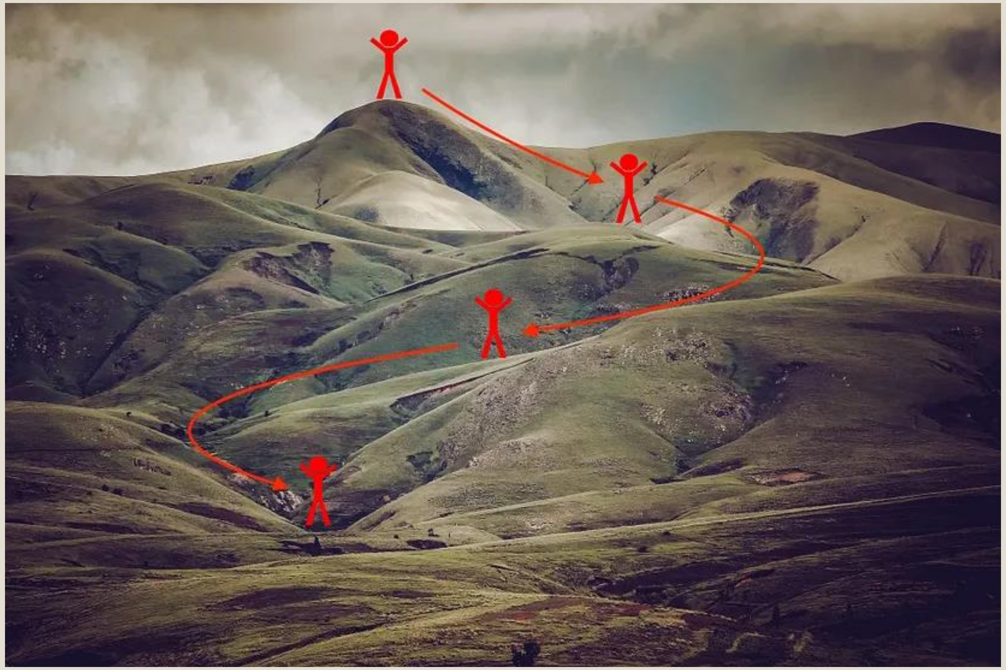


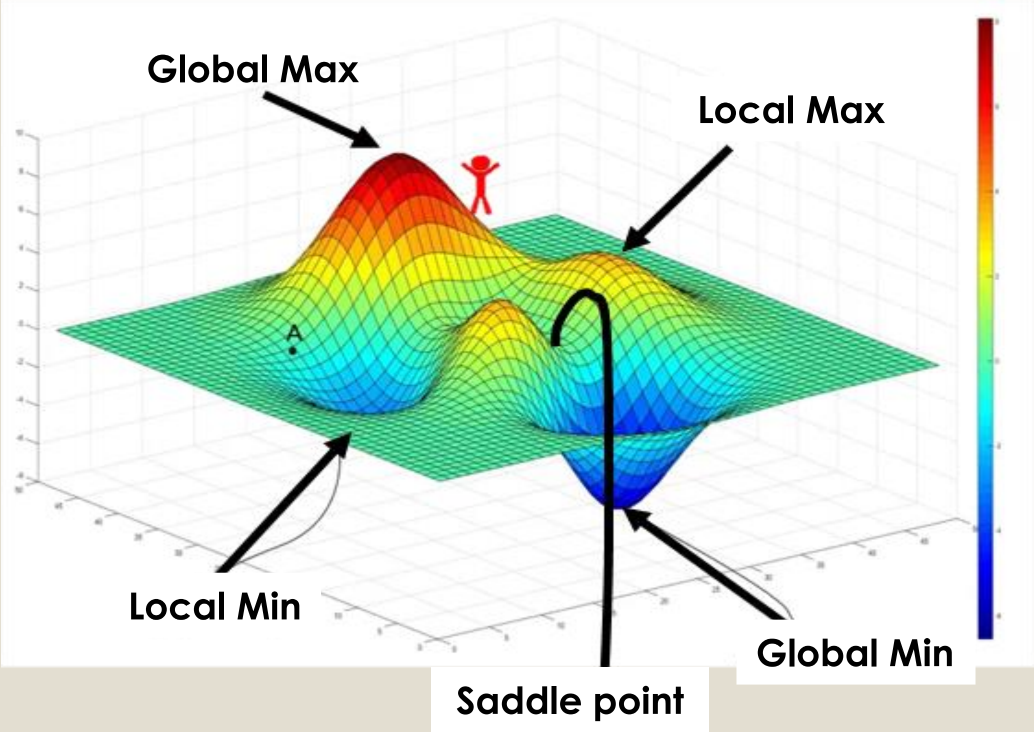


A) 1→ local min, 2 →local max, 3 → saddle point
B) 2→ local min, 1 → local max, 3 → saddle point
C) 3→ local min, 2 → local max, 1 → saddle point
C) 2→ local min, 3 → local max, 1 → saddle point

A) 1→ $f''(x) < 0$, 2 → $f''(x) > 0$, 3 →$f''(x) = 0$
B) 1→ $f''(x) > 0$, 2 → $f''(x) < 0$, 3 →$f''(x) = 0$
C) 1→ $f''(x) = 0$, 2 → $f''(x) < 0$, 3 →$f''(x) > 0$
D) 1→ $f''(x) = 0$, 2 → $f''(x) > 0$, 3 →$f''(x) < 0$

Unconstraint optimization problem-directional search



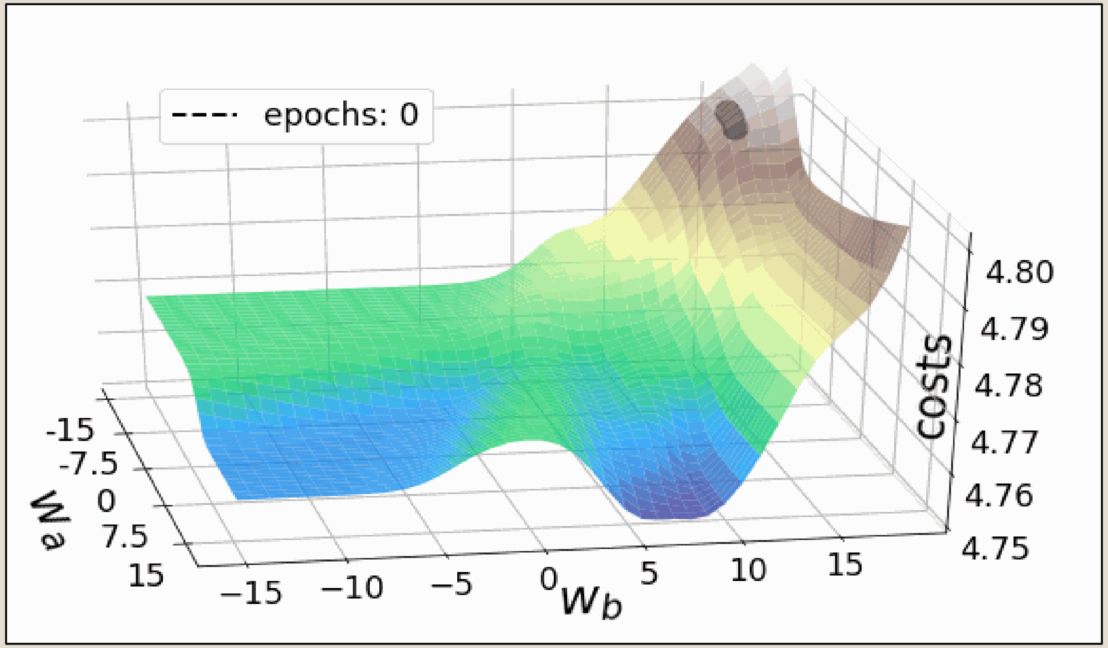


- Aim to reach the bottom point in the region
- Direction of descent
- Steepest descent
- Sometimes we climb up again for better idea to move downwards

Gradient descent method

- Applicable to find the minima of the given function
- Applied in the backpropagation algorithm to find the parameter

$$\boxed{x^{k+1} = x^k + \alpha^k s^k}$$



- Step 1: iteration start at x^k (starting point)
- Step 2: search direction (steepest descent direction)

$s^k = negative\ of\ the\ gradient\ of\ f(x) = -\nabla f$

- Step 3: new point $\boldsymbol{x^{k+1} = x^k + \alpha^k s^k}$ where α^k is the step size (tells how much to move)

Gradient descent illustration 1

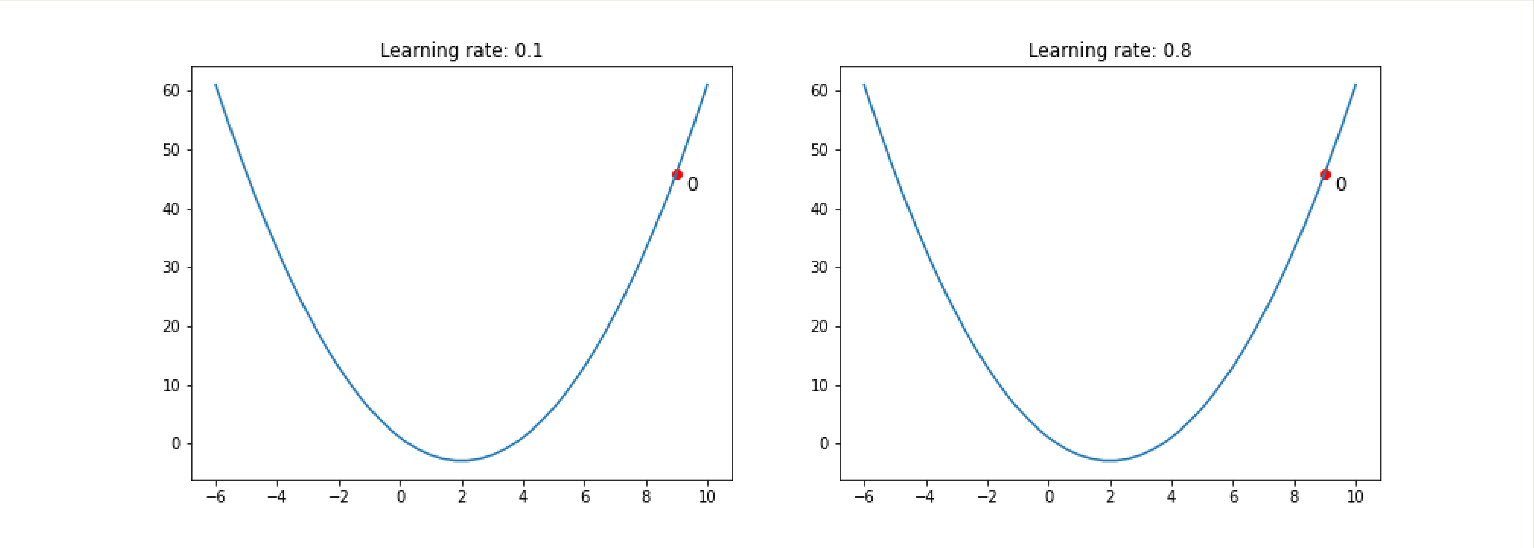
➢Lets understand this using quadratic equation $f(x) = x^2 - 4x + 1$

➢Given learning rate = 0.1, initial point = 9

➢Step 1) calculate the gradient of f(x)→$f'^{(x)} = 2x - 4$

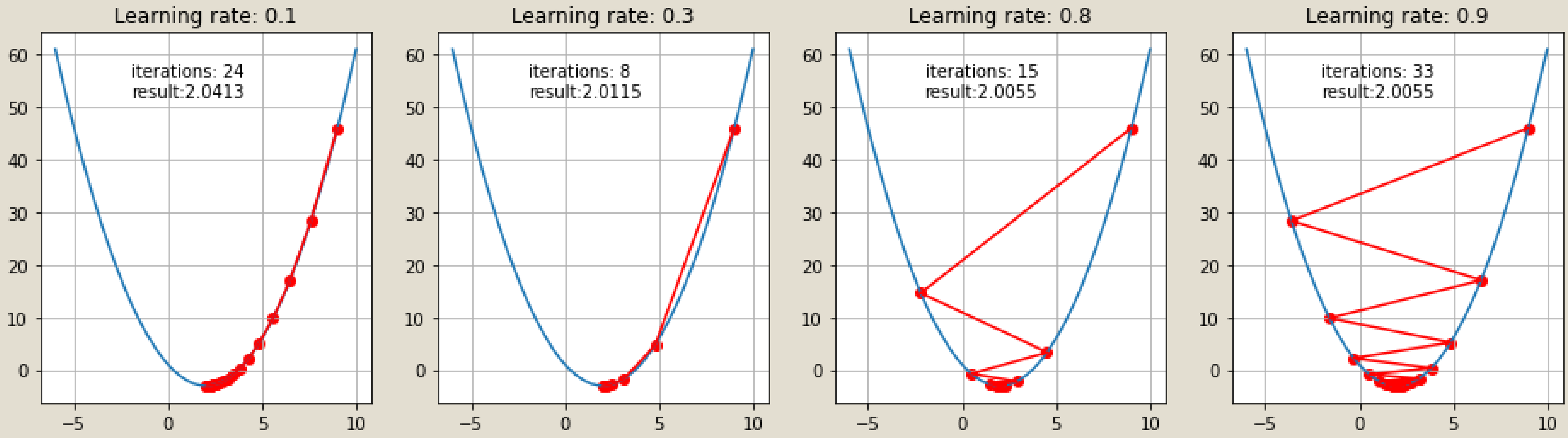
➢Step 2) calculate the next optimal point using $x^{k+1} = x^k - \alpha\, f'(x^k)$

- $x_1 = 9 - 0.1 * (2 * 9 - 4) = 7.6$
- $x_2 = 7.6 - 0.1 * (2 * 7.6 - 4) = 6.48$
- $x_3 = 6.48 - 0.1 * (2 * 6.48 - 4) = 5.584$

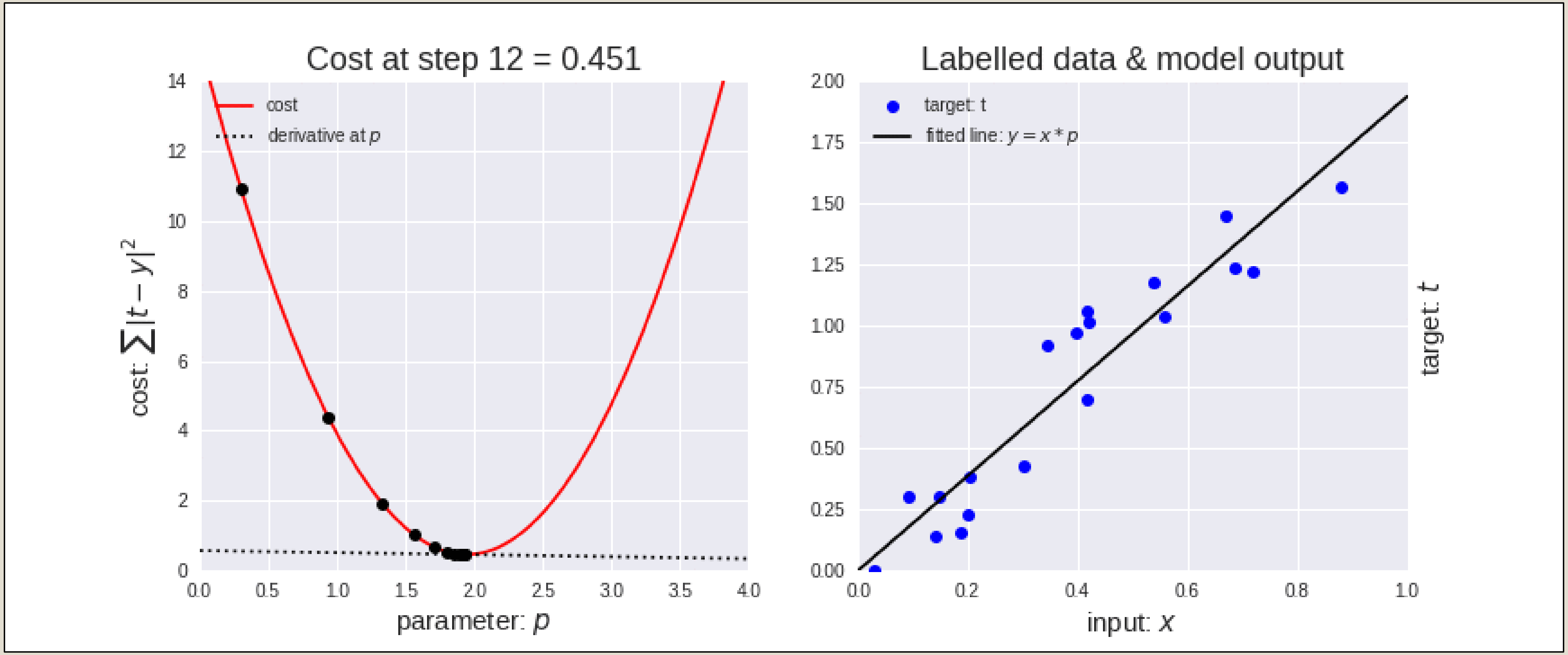


Gradient descent illustration 1

- If learning rate is changed



Gradient descent illustration 2



R Studio

Que: find the values of minima found after 3 iteration for function $f(x) = {x_1}^2 - 2x_1x_2 + 2{x_2}^2 + 2x_1$ with constant step size of 0.5 and initial value (0,0) ?

- The function is $f(x_1, x_2) = {x_1}^2 - 2x_1x_2 + 2{x_2}^2 + 2x_1$
- Finding $\nabla f = \begin{bmatrix} 2x_1 - 2x_2 + 2 \\ -2x_1 + 4x_2 \end{bmatrix}$
- The initial values are x^k=(0,0) and step size α=0.5

$$x^1 = x^0 + \alpha s^0 = x^0 - \alpha \nabla f = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 2(0) - 2(0) + 2 \\ -2(0) + 4(0) \end{bmatrix} = \begin{bmatrix} \mathbf{-1} \\ \mathbf{0} \end{bmatrix}$$

$$x^2 = x^1 + \alpha s^1 = x^1 - \alpha \nabla f = \begin{bmatrix} -1 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 2(-1) - 2(0) + 2 \\ -2(-1) + 4(0) \end{bmatrix} = \begin{bmatrix} \mathbf{-1} \\ \mathbf{-1} \end{bmatrix}$$

$$x^3 = x^2 + \alpha s^2 = x^2 - \alpha \nabla f = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - 0.5 \begin{bmatrix} 2(-1) - 2(-1) + 2 \\ -2(-1) + 4(-1) \end{bmatrix} = \begin{bmatrix} \mathbf{-2} \\ \mathbf{-1} \end{bmatrix}$$

Q) What is the primary goal of the gradient descent algorithm?

A) To find the global minimum of a function
B) To find the critical points of a function
C) To find the local maximum of a function
D) To find the inflection points of a function

Explanation:

The primary goal of the gradient descent algorithm is to find the global minimum (or a local minimum) of a given function.

Q) In the gradient descent algorithm, what is the gradient?
A) A scalar value
B) A vector pointing in the direction of steepest ascent
C) A vector pointing in the direction of steepest descent
D) The second derivative of the function

Explanation:

The gradient is a vector that points in the direction of steepest descent (negative gradient) and indicates the direction in which the function decreases the fastest.

Q) What happens if the learning rate in gradient descent is too small?
A) The algorithm converges quickly
B) The algorithm converges slowly
C) The algorithm oscillates around the minimum
D) The algorithm diverges

Explanation:

A small learning rate causes the algorithm to take small steps in each iteration, leading to slow convergence and a longer time to reach the minimum.

Q) What is the relationship between the learning rate and the risk of overshooting the minimum?
A) Higher learning rate increases the risk of overshooting
B) Lower learning rate increases the risk of overshooting
C) Learning rate has no impact on overshooting
D) Overshooting is unrelated to the learning rate

Explanation:

A higher learning rate can cause the algorithm to take larger steps, which increases the risk of overshooting the minimum and oscillations around it.

Q) What is a disadvantage of using Stochastic Gradient Descent (SGD)?
A) It converges very slowly
B) It requires more memory compared to other methods
C) It can get stuck in local minima
D) It introduces high computational complexity

Explanation:

SGD's random updates can cause it to move in erratic directions and potentially get stuck in local minima instead of finding the global minimum.

Q) In the context of gradient descent, what is the formula for updating the parameter x using the learning rate a and the gradient g?

A) $x_{new} = x - a * g$

B) $x_{new} = x + a * g$

C) $x_{new} = x - g/a$

D) $x_{new} = x + g/a$

Explanation:

n the gradient descent update formula, the parameter x is updated by subtracting the product of the learning rate a and the gradient g from its current value.

Q) f the gradient of a function is [−2,3] and the learning rate is 0.1, what is the updated gradient descent step?

A) [0.2,- 0.3]

B) [−0.2, −0.3]

C) [-0.2, 0.3]

D) [0.2, 0.3]

Explanation:

The updated step is calculated by subtracting *a* times the gradient from the current values: [-2*0.1, 3*0.1] = [-0.2, 0.3]

Q) If the initial parameter value is 8 and the gradient descent step is 0.2, what is the updated parameter value after one iteration?

A) 7.8

B) 8.2

C) 8.0

D) 8.4

Explanation:

The updated parameter value is calculated by subtracting the gradient descent step from the initial value: 8-0.2=7.8