Week-4

Data Science for Engineers (NPTEL)

TA:

Saladi Pravallika,

Prime Ministers Research Fellow

PhD Scholar, IISc Bangalore.

- → Session is being recorded
- → 3-5 min break every half hour & 10 min break every hour.
- → If you would like me to cover more topics in next session please let me know at the end.
- → Be respectful during the session, I will not hesitate to remove you from the session

@ Geeks for Geeks

Optimization

Example : Maps

1. Optimal time path — A
2. Optimal distance path — B
3. Other route — C

→ Optimization is use of a specific method to determine the best possible solution.

Univariate Optimization Problems

1. Objective Function
2. Decision variable
3. Constraints

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

$$x \in R$$

1. Global Maxima — 2

2. Global Minima — 3

3. Local Maxima — 4 — (x_1, x_2)

4. Local Minima — 5 — (x_1, x_2)

* Condition for x to be a minimizer.

$$f'(x) = 0$$ — neccessary condition

Ex: $f(x) = x \sin(x^2) + 1$

Sufficient condition:

$$\boxed{f''(x) > 0}$$

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

(a) find stationary points

(b) find values of x at which minima exist

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

$$f'(x) = \cancel{4} \times \frac{x^3}{\cancel{4}} - \cancel{3} \times \frac{x^2}{\cancel{3}} - 2x$$

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

$$f'(x) = 0$$

$$x^3 - x^2 - 2x = 0$$

$$x(x^2 - x - 2) = 0 \longrightarrow x = 0$$

$$\downarrow$$

$$x^2 - x - 2 = 0$$

$$x^2 - 2x + x - 2 = 0$$

$$(x-2)(x+1) = 0$$

$$x = -1, 2$$

$$\boxed{x = -1, 0, 2}$$

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

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

Put stationary values in the above eq^n $f''(x)$

$x = -1, 0, 2$

$$f''(-1) = 3(-1)^2 - 2(-1) - 2$$
$$= 3(1) + 2(1) - 2$$
$$= 3 + 2 - 2 = 3 //.$$

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

$$f''(2) = 3(2)^2 - 2(2) - 2$$
$$= 3(4) - 4 - 2$$
$$= 12 - 6 = 6 //.$$

$f''(x) > 0$

$$\left.\begin{aligned} f''(-1) &= 3 \\ f''(2) &= 6 \end{aligned}\right\} \longrightarrow \text{Minima exists } \underline{\text{at } x = +2}$$

Substitute $x = +2$ in $f(x)$

Ques:

What is the minimum point of 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) $x = -1$, $f(x) = 2$

$$f'(x) = 2x - 4$$

$$f'(x) = 0$$

$$2x - 4 = 0$$

$$2x = 4$$

$$\boxed{x = 2}$$

$$f(2) = 2^2 - 4(2) + 5$$

$$= 4 - 8 + 5$$

$$= -4 + 5 = 1$$

$$\boxed{f(2) = 1}$$

Question:

Of all possible rectangles with area = 100, which has smallest perimeter.

Hint: Area = $l \times b$

perimeter = $2l + 2b$

$A = 100$

$l \times b = A$ $\qquad 2l + 2b$

$b = \frac{A}{l} \rightarrow (1) \rightarrow 2l + 2\left(\frac{A}{l}\right)$

$= 2l + \frac{2(100)}{l}$

$= 2l + \frac{200}{l}$.

$f(x) = 2l + \frac{200}{l}$

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

$\frac{200}{l^2} = 2$

$l^2 = \frac{200}{2} = 100$

$l = +10, -10$

l cannot be -ve, $l = +10$.

$b = \frac{A}{l} = \frac{100}{10} = 10$

$P = 2l + 2b = 2(10) + 2(10) = 40$ //.

Question: DIY

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 ?

Answer : 500 units

R-Studio

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

$P = 2l + 2b = 24 \Rightarrow l + b = 12$

Max Area.

$l = 12 - b \rightarrow (1)$

$A = l \times b$

$= (12-b)\,b$

$= 12b - b^2$

$A' = 12 - 2b = 0$

$\boxed{b = 6}$

$l = 12 - b$

$= 12 - 6 \Rightarrow \boxed{l = 6}$

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) $x = 1, x = 3$

$$f'(x) = 3x^2 - 12x + 9 = 0$$

$$3(x^2 - 4x + 3) = 0$$

$$x^2 - 3x - x + 3 = 0$$

$$(x-3)(x-1) = 0$$

$$\boxed{x = 1, 3}$$

A company wants to maximize its profit P

$$P(x) = -2x^2 + 40x + 100$$

, where x is
the quality of products sold. What is
optimal ~~quantity~~ quality of product to sell?

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

$$f'(x) = -4x + 40 = 0$$

$$\boxed{x = 10}$$

A car travels along a straight road. Its
position is given by

$$s(t) = 2t^2 + 3t + 5$$

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

Velocity → derivative of dist.

Velocity → $s'(t)$

$$s'(t) = 4t + 3$$

$t = 2$

$$s'(2) = 4(2) + 3$$
$$= 8 + 3$$
$$= 11 \text{ m/sec}$$

Multi-variate Optimization

$z = f(x_1, x_2, \dots, x_n)$

let $z = \sqrt{x^2 + y^2}$

local minima = (0, 0)

Some level sets of $z = x^2 + y^2$

$C = x^2 + y^2$

$\longrightarrow$ Contour Plot

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) $f''(x) < 0$

Univariate

Multi-variate Optimization condition.

$$\boxed{z = f(x_1, x_2 \cdots, x_n)}$$

(i) Necessary condition $\longrightarrow$

$$\nabla f(x^*) = 0$$

$\hookrightarrow$ Gradient $= \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \partial f / \partial x_3 \\ \vdots \end{bmatrix}$

$f(x) = 2x_1 + 3x_2$

$\frac{\partial f(x)}{\partial x_1} = 2$

$\frac{\partial f(x)}{\partial x_2} = 3$

(ii) Sufficient Condition

$\nabla^2 f(x^*) \rightarrow$ shld have +ve definite

$\hookrightarrow$ Hessian $\rightarrow$

$\downarrow$

matrix with +ve eigen values

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

$$\begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ \cdots & - & \cdots \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \partial^2 f / \partial x_1 \cdot \partial x_1 \; (\partial x_1^2) & \partial^2 f / \partial x_1 \partial x_2 & \partial^2 f / \partial x_1 x_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{Hessian}}$$

Eigen values are +ve indicates that matrix is +ve definite & represents existence of local minima.

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

$$f(x_1, x_2) = x_1 + 2x_2 + 4x_1^2 - x_1x_2 + 2x_2^2$$

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

$$\frac{\partial f}{\partial x_1} = 1 + 0 + 8x_1 - x_2 + 0$$

$$= 8x_1 - x_2 + 1 \longrightarrow (1)$$

$$\frac{\partial f}{\partial x_2} = 0 + 2 + 0 - x_1 + 4x_2$$

$$= -x_1 + 4x_2 + 2 \longrightarrow (2)$$

$$8x_1 - x_2 + 1 = 0$$

$$-x_1 + 4x_2 + 2 = 0$$

$$x_1 = -0.19$$

$$x_2 = -0.54$$

$$\nabla^2 f = \begin{bmatrix} \partial f / \partial x_1^2 & \partial f / \partial x_1 \partial x_2 \\ \partial f / \partial x_1 \partial x_2 & \partial f / \partial x_2^2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & -1 \\ -1 & 4 \end{bmatrix}$$

Condition: eigen values are +ve.

R-Studio

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

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

Question

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

$$\frac{\partial f}{\partial x} = 8x - 3y \xrightarrow{x=0,\, y=1} 8(0) - 3(1) = -3$$

$$\frac{\partial f}{\partial y} = -3x + 12y \xrightarrow{x=0,\, y=1} -3(0) + 12(1) = 12.$$

$$[-3, 12]$$

Question

Calculate Hessian Matrix for $f(x,y) = 4x^2 - 3xy + 6y^2$ at pt $(0,1)$

$$\frac{\partial f}{\partial x} = 8x - 3y \Rightarrow \frac{\partial^2 f}{\partial x_1^2} = 8$$

$$\frac{\partial f}{\partial y} = -3x + 12y \qquad \frac{\partial^2 f}{\partial x \partial y} = -3$$

$$\frac{\partial^2 f}{\partial y^2} = 12$$

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

Question

A f^n $f(x,y) = x^2 + 4y^2$ has critical point at $(2,1)$.

What is the value of f^n at this critical point?

$$f(2,1) = 2^2 + 4(1)$$
$$= 4 + 4 = 8$$

Question

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

What can be said about critical point?

$\rightarrow$ $\nabla^2 f = \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$

eigen values – $5 \pm \sqrt{5}$

Both are +ve

$\rightarrow$ It is local minima.

Unconstrained 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

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

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

$\alpha^k \rightarrow$ step size

$s^k \rightarrow -\nabla f$

Method

Step 1 : iteration starts at x^k

Step 2 : Search direction $\equiv$ steepest descent direction using s^k $(= -\nabla f)$

Step 3 : New point x^{k+1} (tells how much to move)

Question:

$$f(x) = x^2 - 4x + 1$$

learning rate $= 0.1 \longrightarrow \alpha$

Initial point $= \underline{9} \longrightarrow x^k$

Hint : Optimal point $= x^{k+1} = x^k - \alpha f'(x^k)$

(1) $f(x) = x^2 - 4x + 1$

$f'(x) = 2x - 4$

$$x^{k+1} = 9 - (0.1)(2(9) - 4)$$
$$= 9 - (0.1)(18 - 4)$$
$$= 9 - (0.1)(14)$$
$$= 9 - 1.4$$
$$= 7.6.$$

(1) $\underset{x^k}{9} \rightarrow \underset{x^{k+1}}{7.6}$

(2) $\underset{x^k}{7.6} \rightarrow \underset{x^{k+1}}{___}$

(3)

⋮

What happens if α is changed?

R- Studio

Question

find values of minima found after 3 iterations

$$f(x) = x_1^2 - 2x_1x_2 + 2x_2^2 + 2x_1$$

const. step size = 0.5
initial value = (0, 0)

find ∇f

$x^k = (0,0) \quad \alpha = 0.5$

$x^1 = x^0 - \alpha \nabla f$ — calculate

$= \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$x_2 = x_1 - \alpha \nabla f$ — calculate

$= \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

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

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

$x^{k+1} = x^k - \alpha \nabla f$

descent

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

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

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

If 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]

x^k

$lr \times [x, y]$

$[-2, 3] \times 0.1$

$[-0.2, 0.3]$

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

$\alpha \nabla f = 0.2$

$x_k = 8$

$x_{k+1} = x_k - \alpha \nabla f$

$= 8 - 0.2$

$= 7.8$