Gradient Descent Algorithm (GDA)

21 September 2025 11:24

* Pro-tip Loss Function vs Cost Function

Loss Function: It is the function to capture the difference between actual and predicted value. for a single row or a single record or a single training example.

$$\text{error}) = [\underset{\text{actual value}}{y_i} - \underset{\text{predicted value}}{\hat{y}_i}] \text{ for some } i^{th} \text{ row}$$

cost Function

actual value Predicted value

X1	X2	X3	X4	X5	X6	X7	Yi

$Y_1 - \hat{Y}_1 = e_1$

$Y_2 - \hat{Y}_2 = e_2$

$Y_3 - \hat{Y}_3 = e_3$

$Y_4 - \hat{Y}_4$

$Y_5 - \hat{Y}_5$ ⋮

$Y_6 - \hat{Y}_6 = e_6$

Regression Model.

[average of errors]

in some way → Cost Function

for regression : MSE → Mean Squared Error

Cost Function:

It is used to refer an average of the loss functions in some way over the entire training dataset

Some famous cost Functions

MSE – Mean Squared Error

MSE : Mean Squared Error

1. error $= (y_i - \hat{y}_i)$

2. squaring the error $= (y_i - \hat{y}_i)^2$

3. Mean of the squared error $= \boxed{\dfrac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$

   $\underbrace{\text{Mean}}_{\text{avg.}}$

$$\boxed{MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$\boxed{RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}$$

$\sqrt{MSE}$

Task: Read about different cost functions for classification problems. → What is a logit???

- Logistic Regression
- Decision Trees
- Support vector Machines
  ↳ SVM...

Binary as well as Multi-class.

Gradient Descent Algorithm (GDA)

Gradient : is the rate of change → differentiation (partial derivatives) or derivative

→ (slope / rate of change)

+

Descent coming down (direction is downwards)

→ Gradient descent tells us how steep the slope is and which direction decreases the error.

[∇J : gradient of J → it points in the direction of maximum increase of the cost/error function]

J ↓ cost function

GD: Gradient Descent moves in the opposite direction of the gradient because we want to minimize the error.

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$$(\text{for } j = 1 \text{ and } j = 0)$$

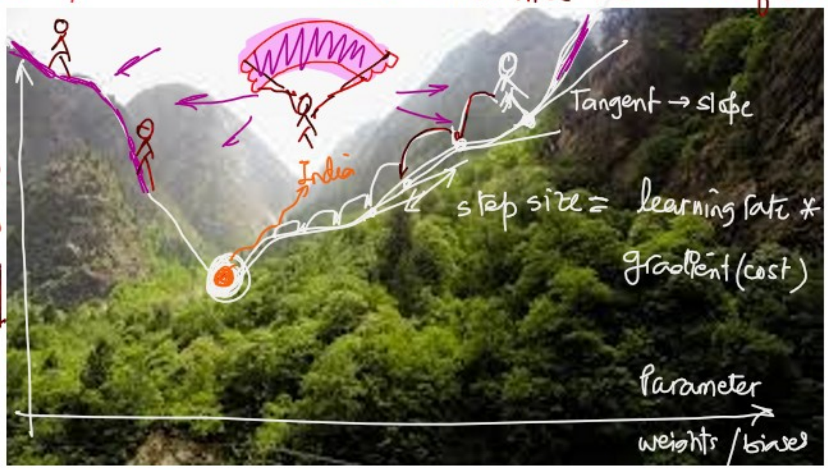
}

Intuition : Abhinandan – Mig-21 Pilot got captured on the enemy's land.

© obj: Abhinandan needs to minimize the chance of getting caught.

© apc: Abhinandan needs to minimize the chance of getting caught.

cost/loss Function
↓
Risk of losing life

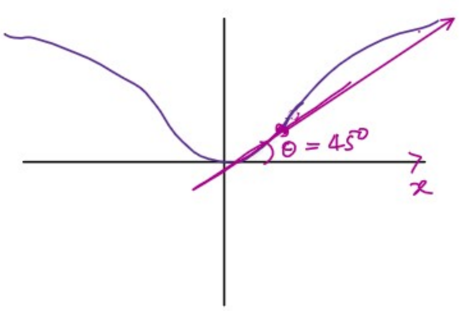


Gradient → slope → $\frac{dy}{dx}$ → $\tan\theta$

Descent → coming down.

Gradient Descent → follows the path of the steepest descent taking steps in the direction that decreases the slope and brings Abhinandan closer to the local minimum (coming to plains)

$$\boxed{\tan\theta = \text{slope} = m = \frac{dy}{dx}}$$

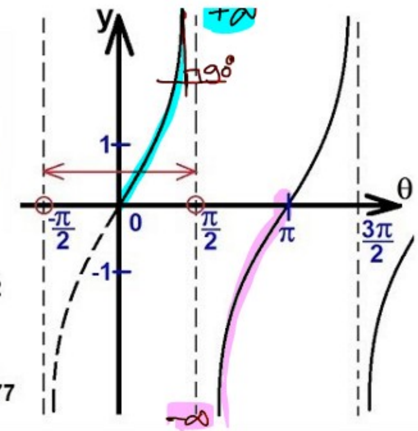


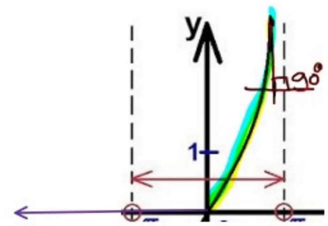
$\tan(\theta) = \tan 45^\circ = 1$
↓
$\text{slope} = \frac{dy}{dx} = m = 1$

Graph for tangent function

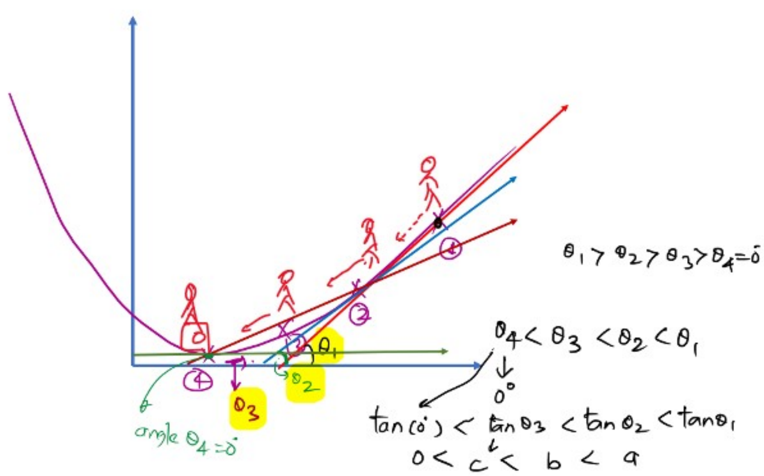
θ (degree)	θ (radian)	tanθ
0	0	0 0

θ (degree)	θ (radian)	tanθ	
0	0	0	0
30	$\pi/6$	$1/\sqrt{3}$	0.577
45	$\pi/4$	1	1
60	$\pi/3$	$\sqrt{3}$	1.732
90	$\pi/2$	∞	∞
120	$2\pi/3$	$-\sqrt{3}$	-1.732
135	$3\pi/4$	-1	-1
150	$5\pi/6$	$-1/\sqrt{3}$	-0.577





[redacted]



① As 'θ' increases from 0° till 90° → tanθ → slope also increases

② and as θ increases from 90° till 180° → tanθ → slope also decreases