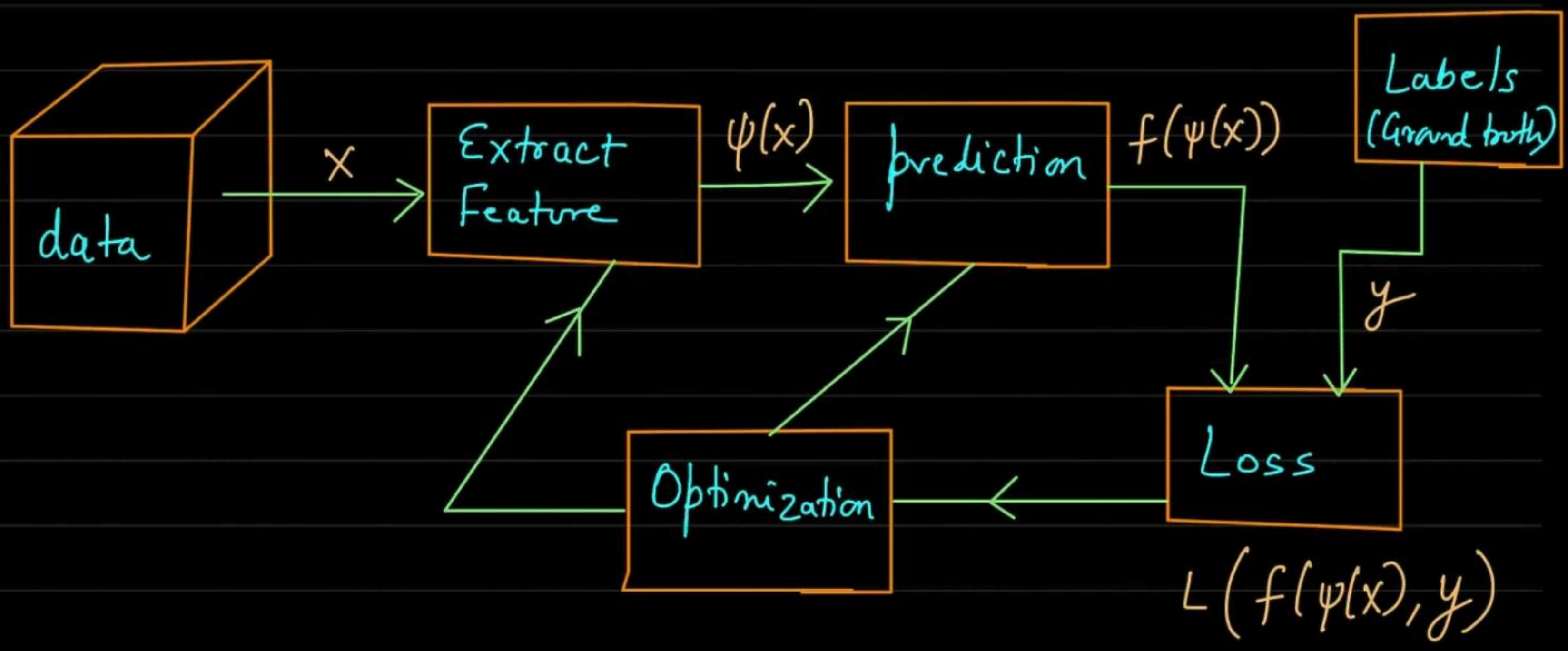
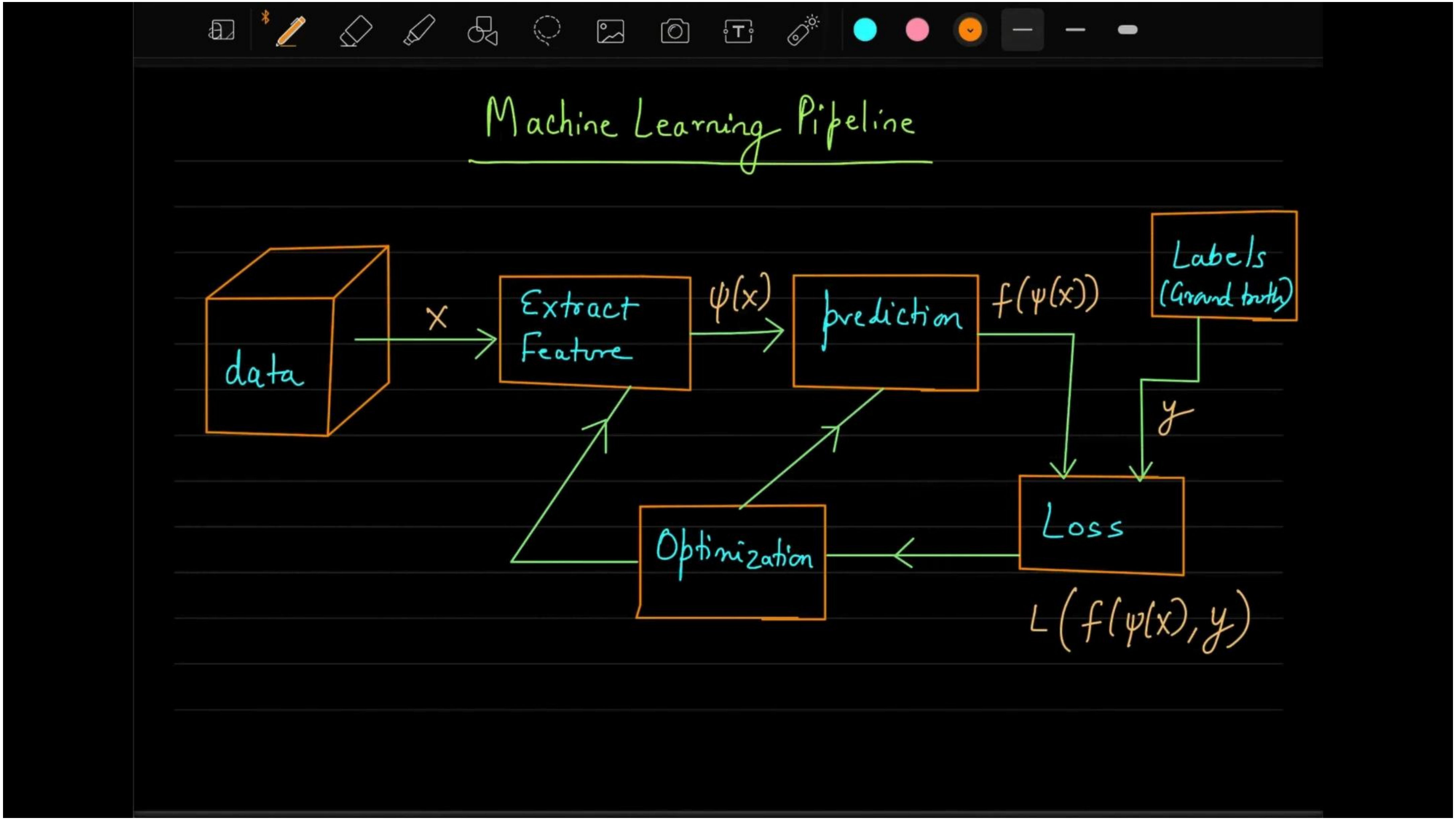
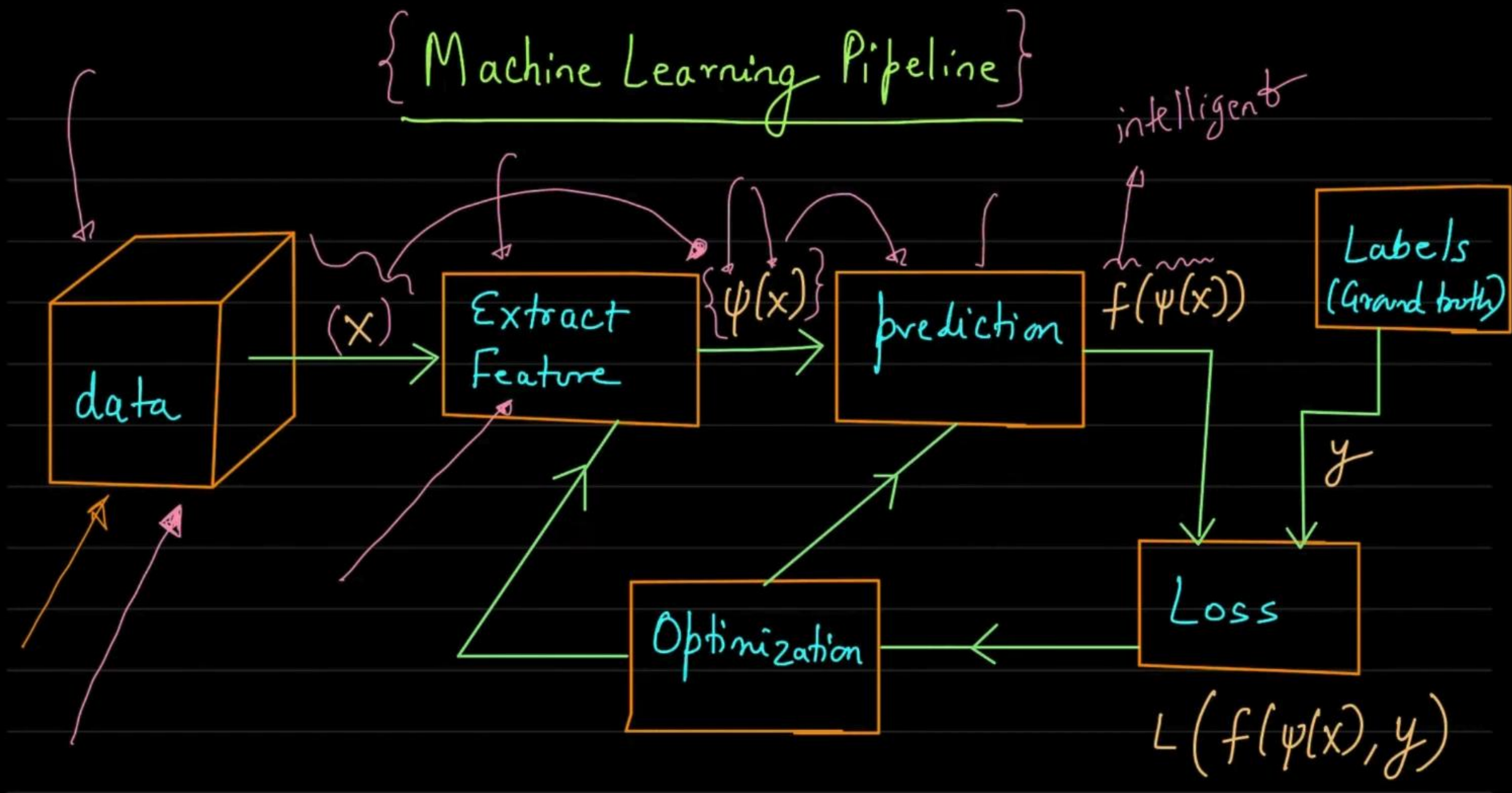
Machine Learning Pipeline



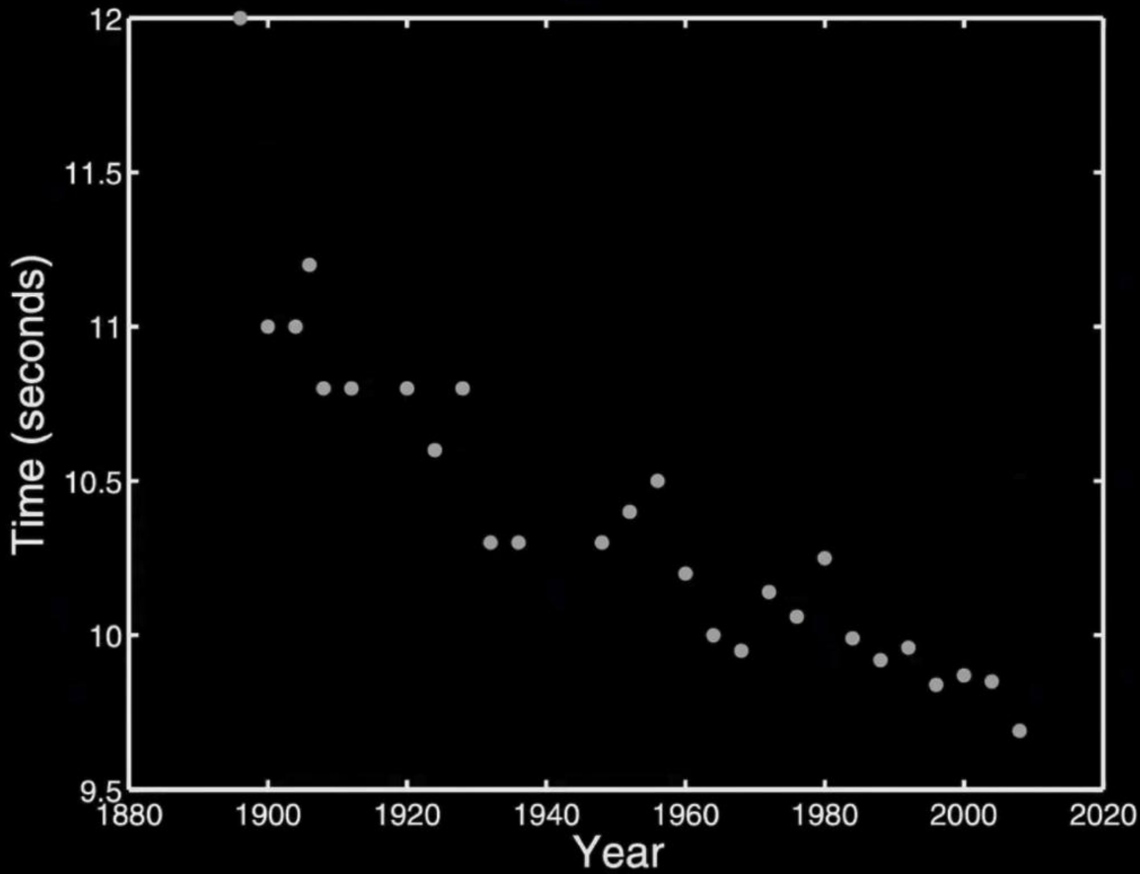

Machine Learning Pipeline
data
x
Extract Feature
ψ(x)
prediction
f(ψ(x))
Labels (Grand truth)
y
Loss
Optimization
L(f(ψ(x), y)



{Machine Learning Pipeline}
intelligent
data
(X)
Extract
Feature
{ψ(X)}
prediction
f(ψ(X))
Labels
(Grand truth)
y
Loss
Optimization
L(f(ψ(X), y)

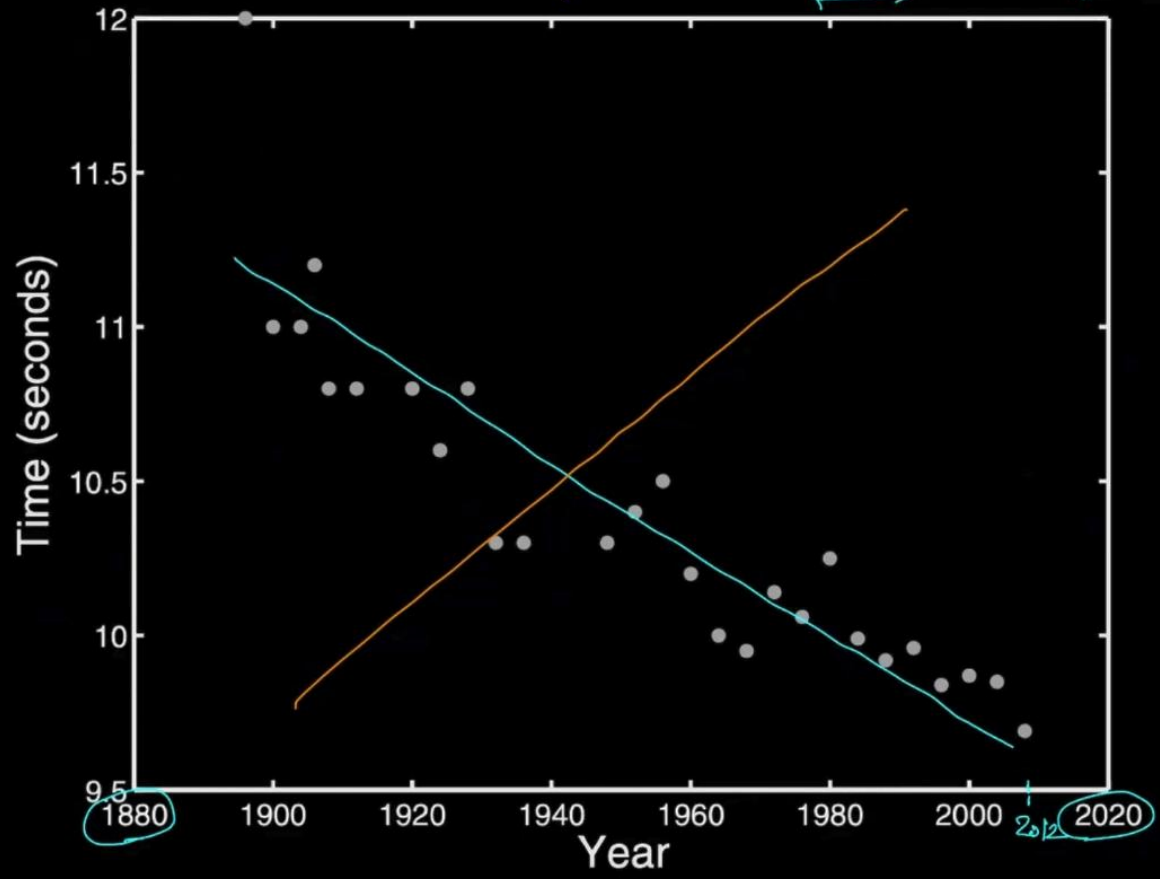
Linear Regression

Winning men's 100m at Summer Olympics



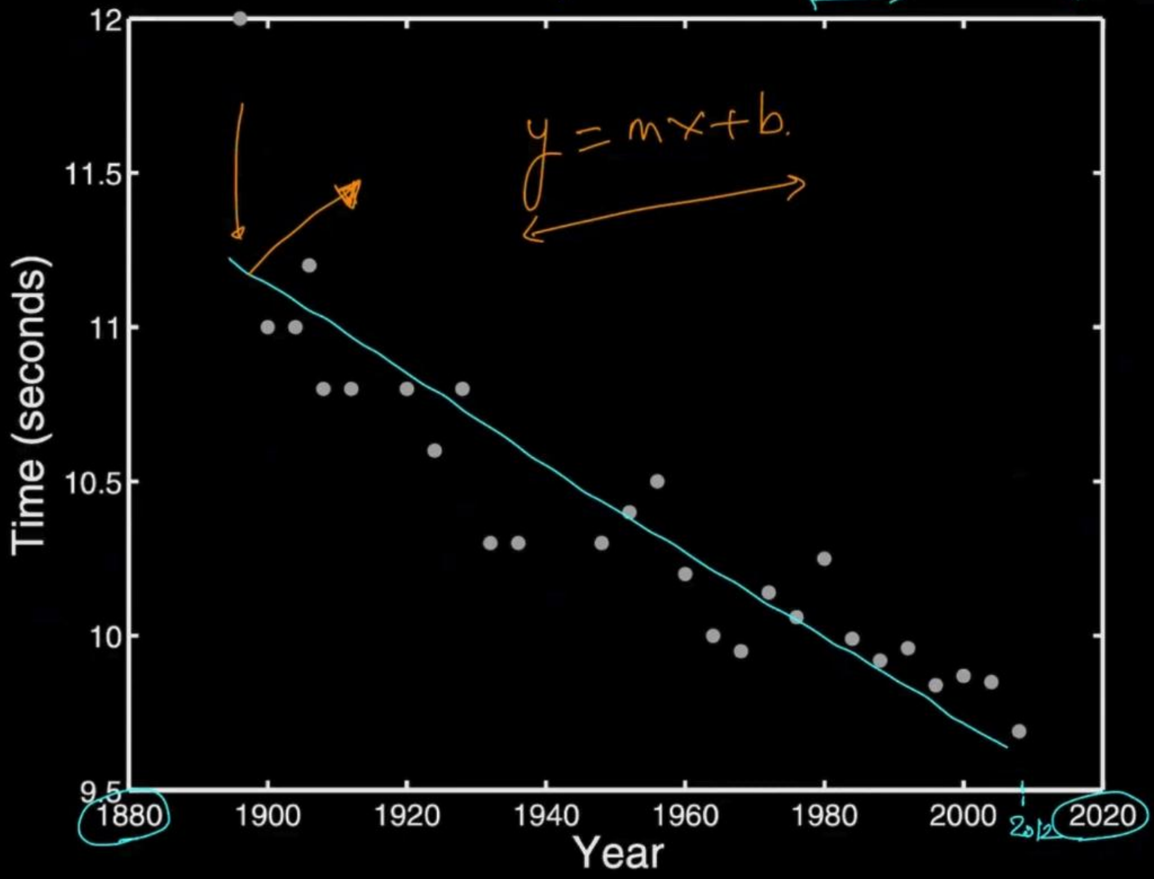
Linear Regression

Winning {men's} 100m at Summer Olympics



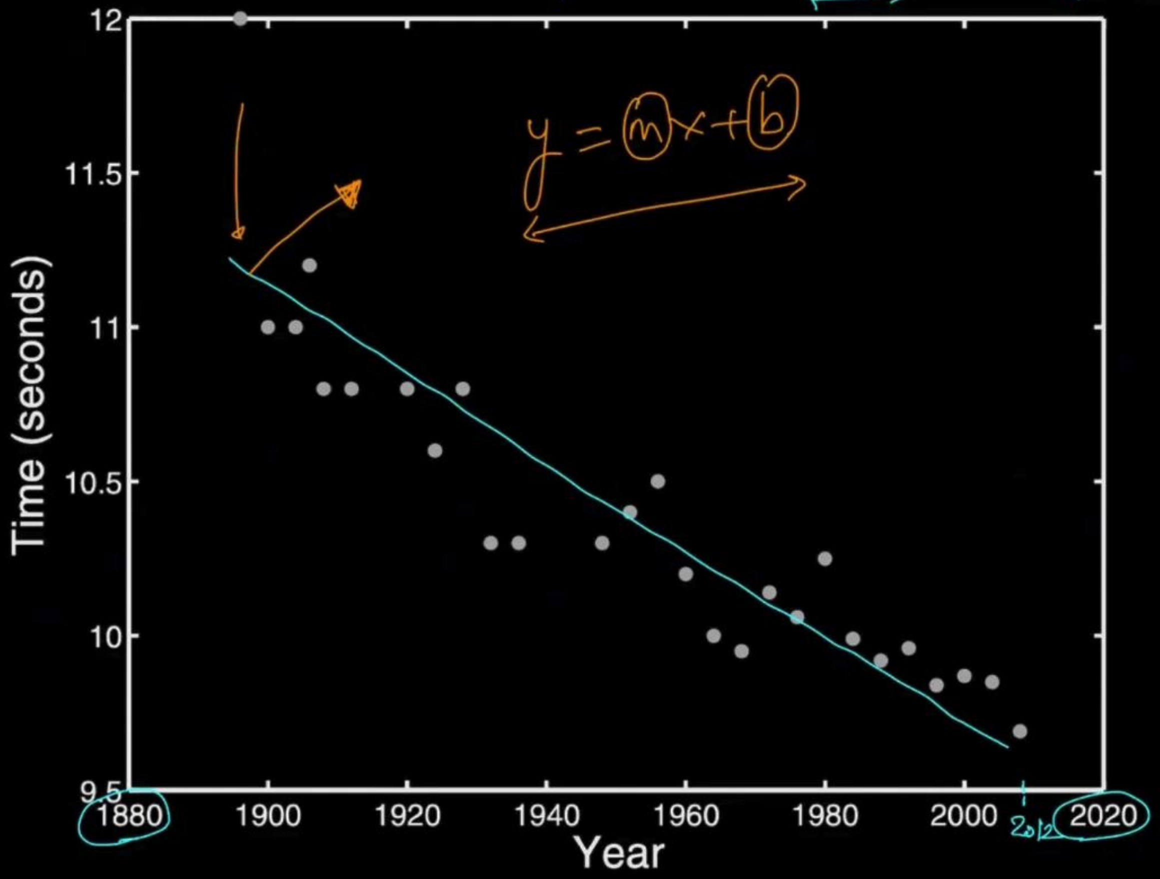
Linear Regression

Winning {men's} 100m at Summer Olympics



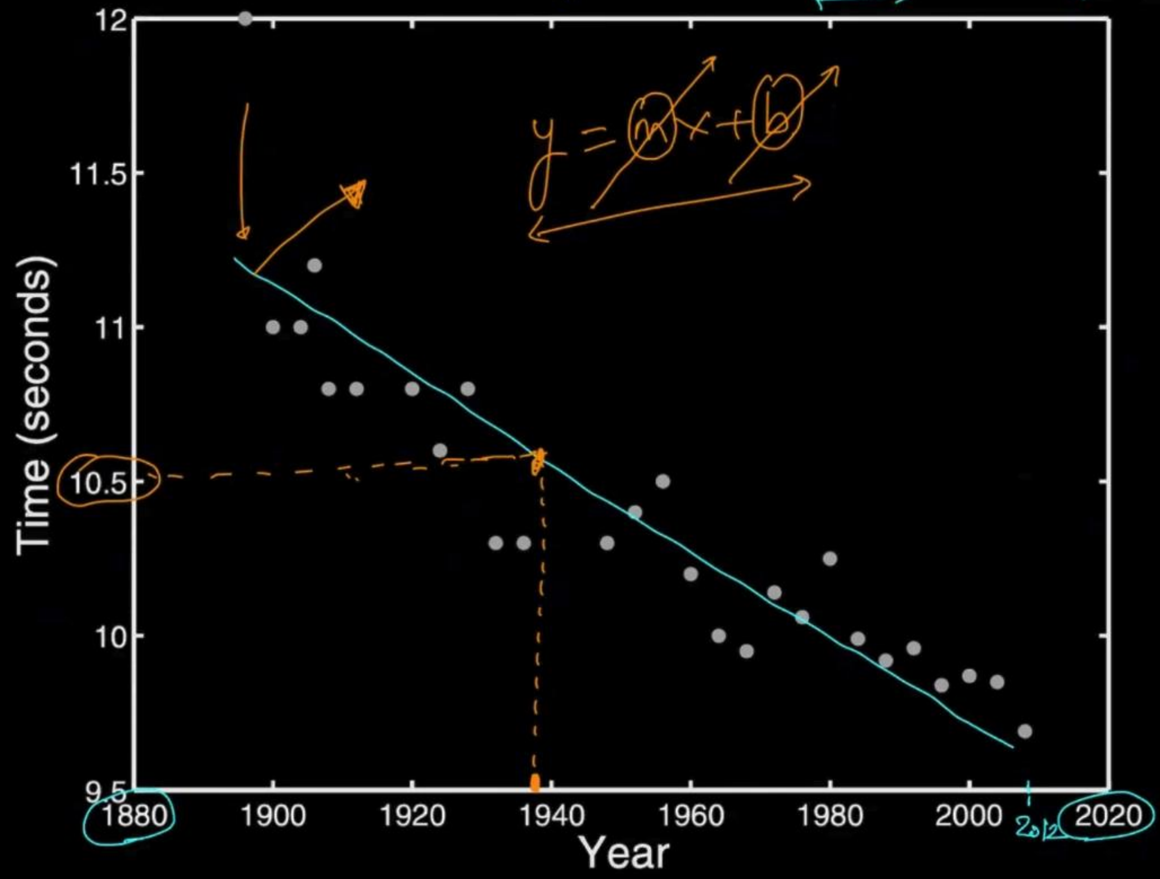
Linear Regression

Winning {men's} 100m at Summer Olympics



Linear Regression

Winning {men's} 100m at Summer Olympics



i	$x_1^{(i)}$	$y^{(i)}$
1	1896	12.00
2	1900	11.00
3	1904	11.00
4	1906	11.20
5	1908	10.80
6	1912	10.80
7	1920	10.80
8	1924	10.60
9	1928	10.80
10	1932	10.30
11	1936	10.30
12	1948	10.30
13	1952	10.40
14	1956	10.50
15	1960	10.20
16	1964	10.00
17	1968	9.95
18	1972	10.14
19	1976	10.06
20	1980	10.25
21	1984	9.99
22	1988	9.92
23	1992	9.96
24	1996	9.84
25	2000	9.87
26	2004	9.85
27	2008	9.69

Mathematical Notation

$i \rightarrow$ index to number of sample

Input : $x_n^{(i)}$ $\leftarrow$ sample (i), $\leftarrow$ feature (n)

Output : $y^{(i)} \longrightarrow$ Labels

Training Data : $\{ x_n^{(i)}, y^{(i)} \}$

For this example

$i = 1$ to 27

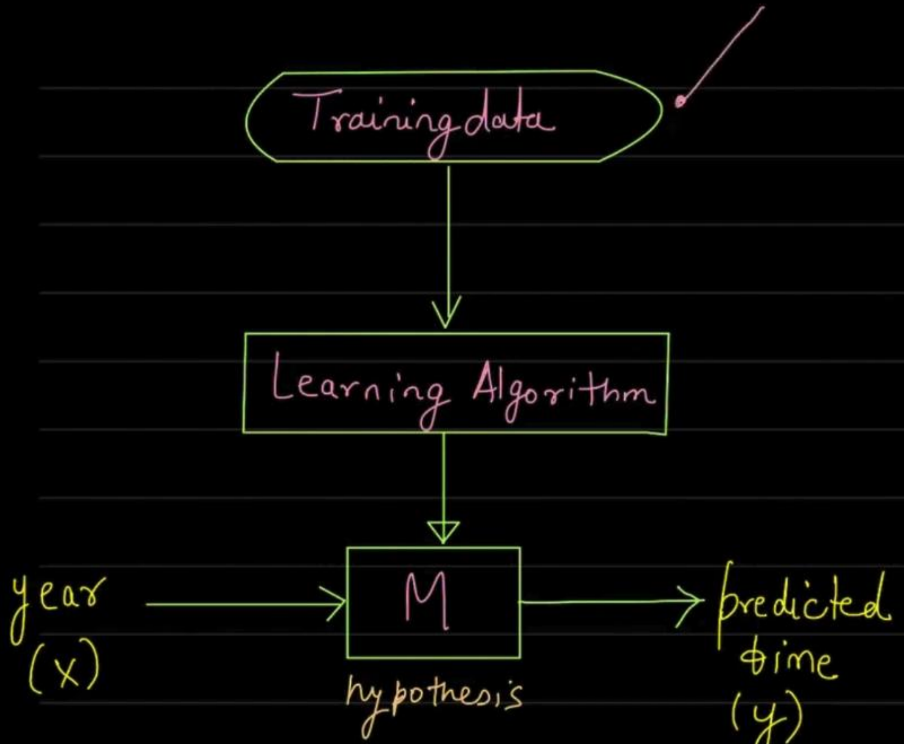
$n = 1$ (one feature)

m : training examples

$m = 27$

$x_1^{(1)} = 1896$, $x_1^{(14)} = 1956$

$y^{(1)} = 12$, $y^{(14)} = 10.50$



M is a function that maps X's to y's

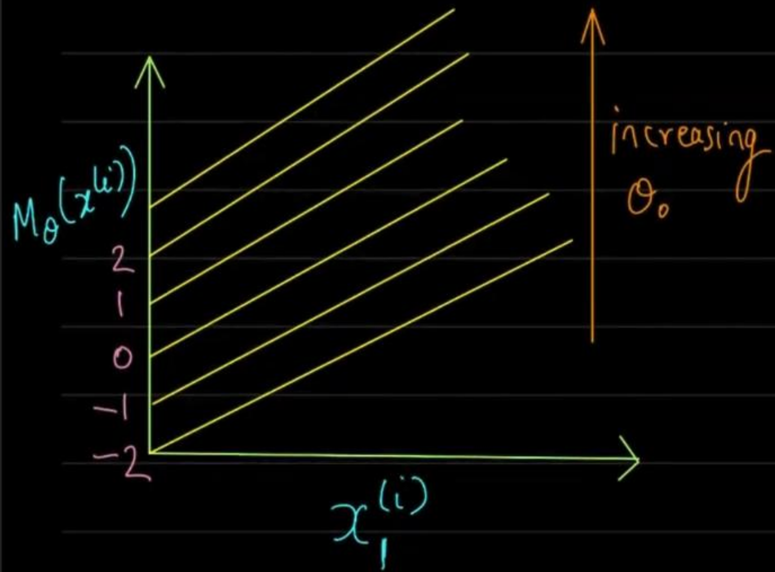
Q. How do we represent hypothesis M ?

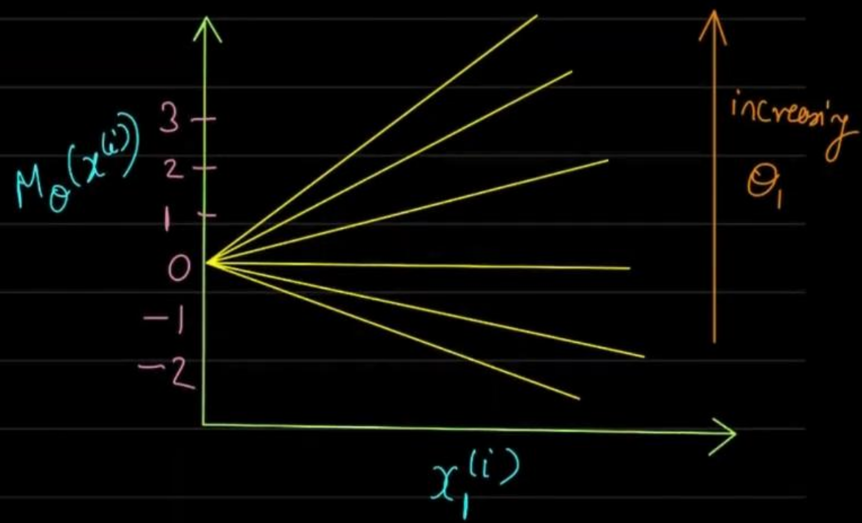
(1

Model Representation for One Variable/Univariate Regression

Parameters

$$M_\theta(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)}$$





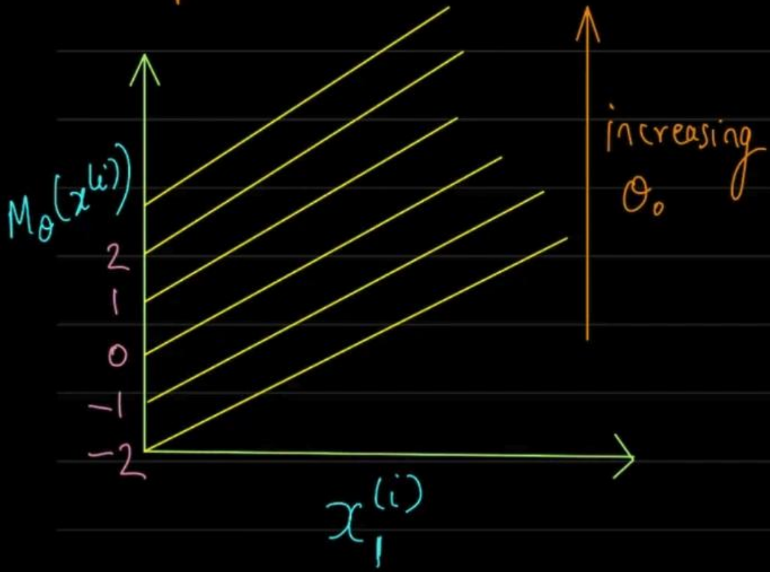
("M")

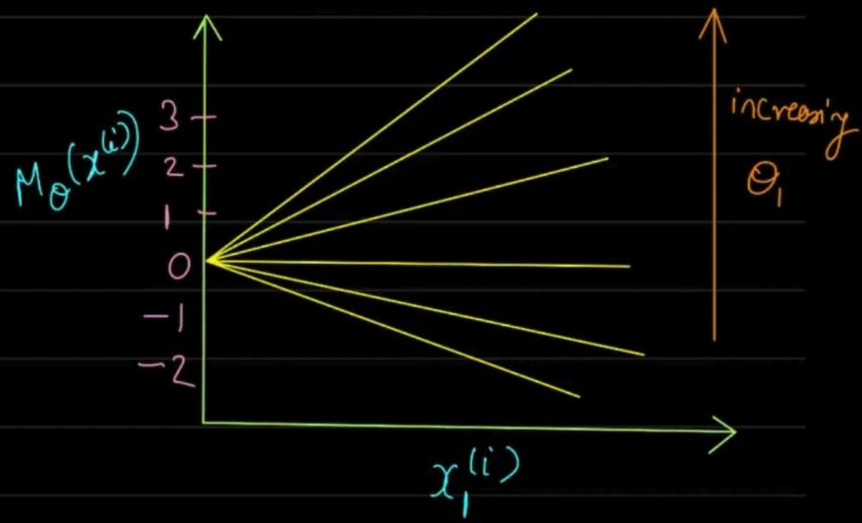
Model Representation for One Variable/Univariate Regression

Parameters

"1 to 27"

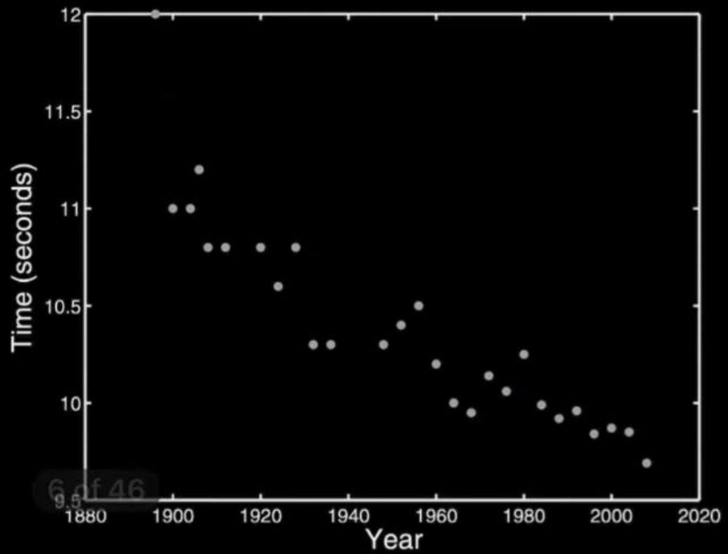
$$M_\theta(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)}$$





Defining a Good Model (Selection of parameters)

How to select optimal value of θ_0 & θ_1 ?



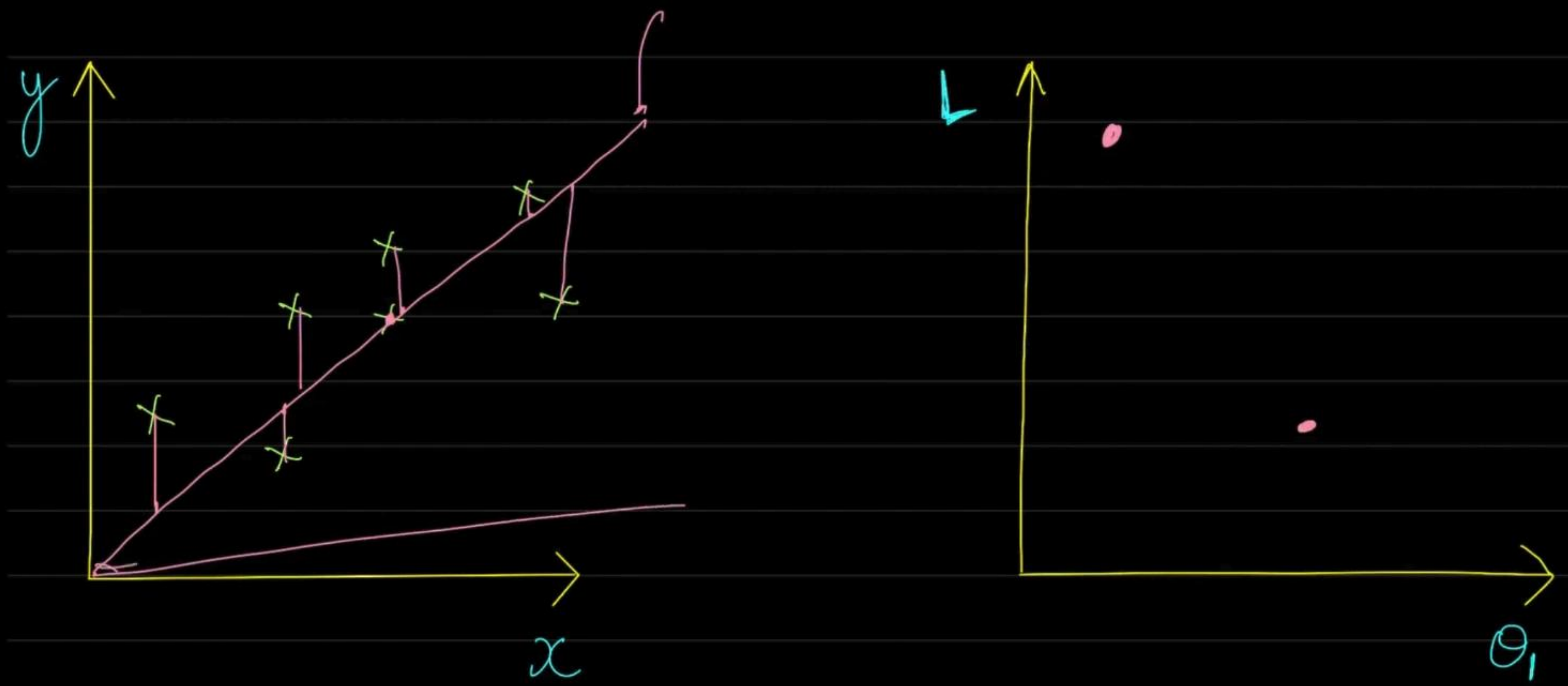
$$\text{Loss}(L) = \underset{\theta_0,\,\theta_1}{\operatorname{argmin}} \frac{1}{2m} \sum_{i=1}^{m} \left(M_\theta\left(x^{(i)}\right) - y^{(i)} \right)^2$$

"Least square Cost function"

Curvature of the Loss function

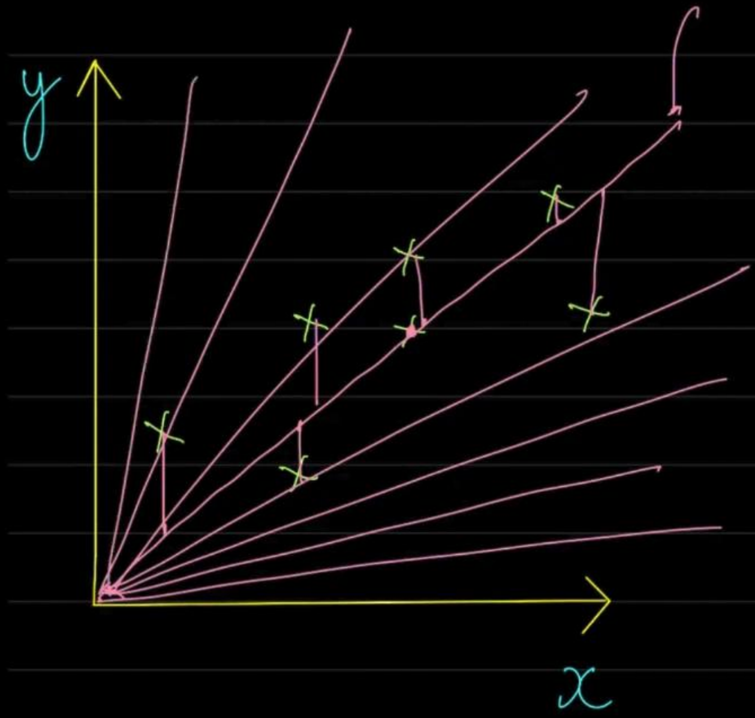
Assume $\theta_0 = 0$

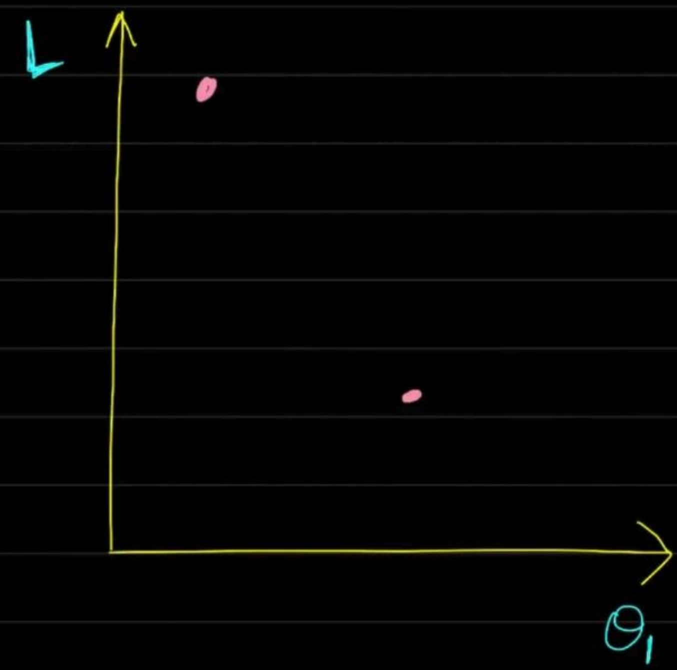
Curvature of the Loss function



Assume $\theta_0 = 0$

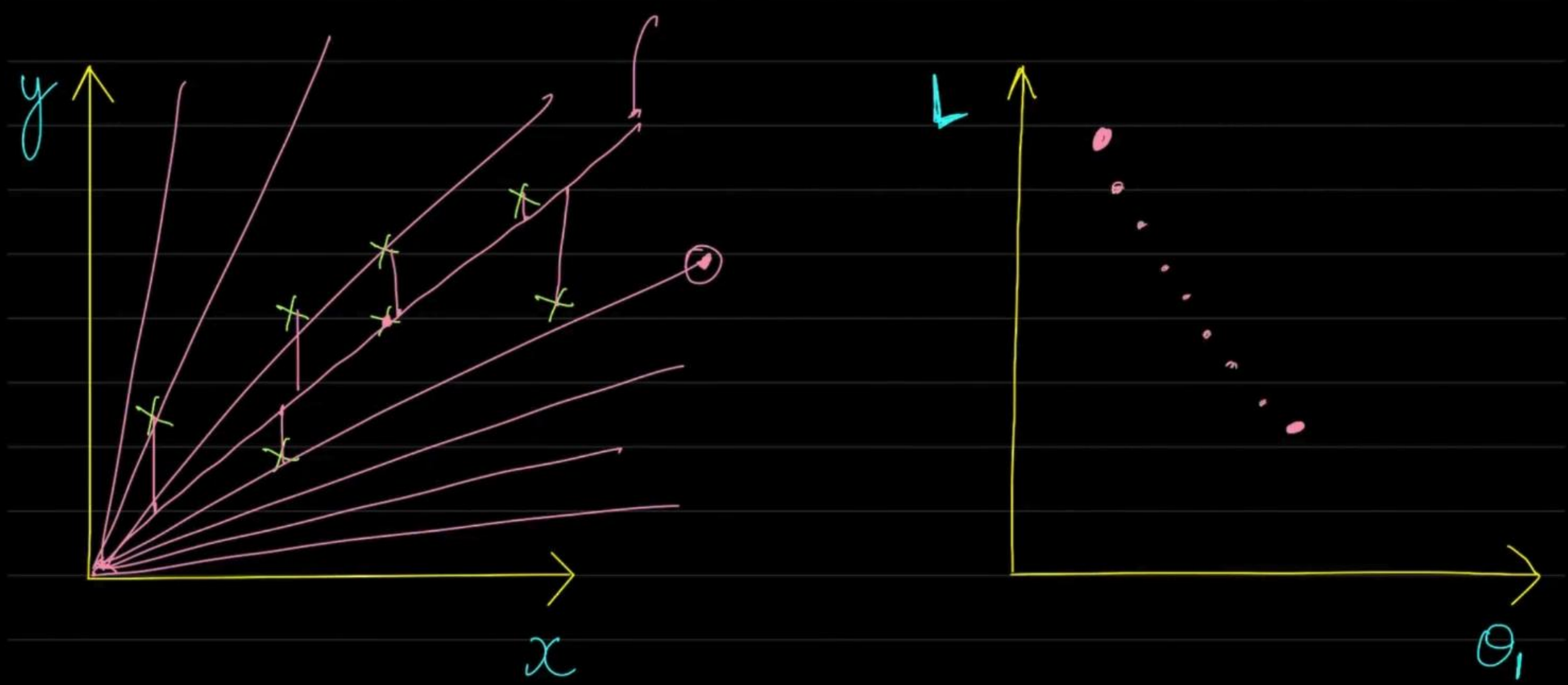
Curvature of the Loss function





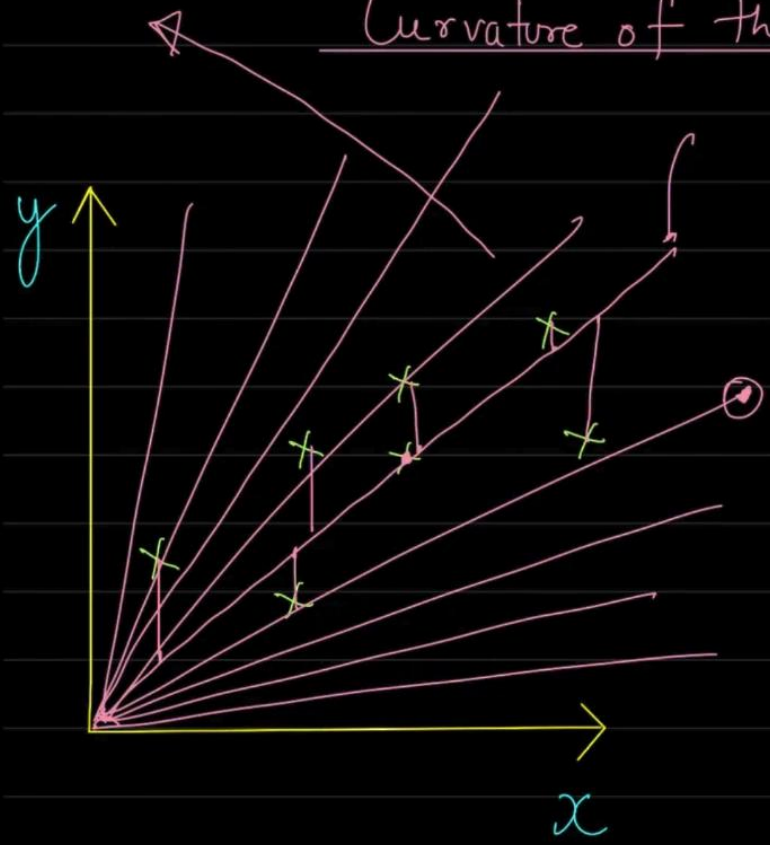
Assume $\theta_0 = 0$

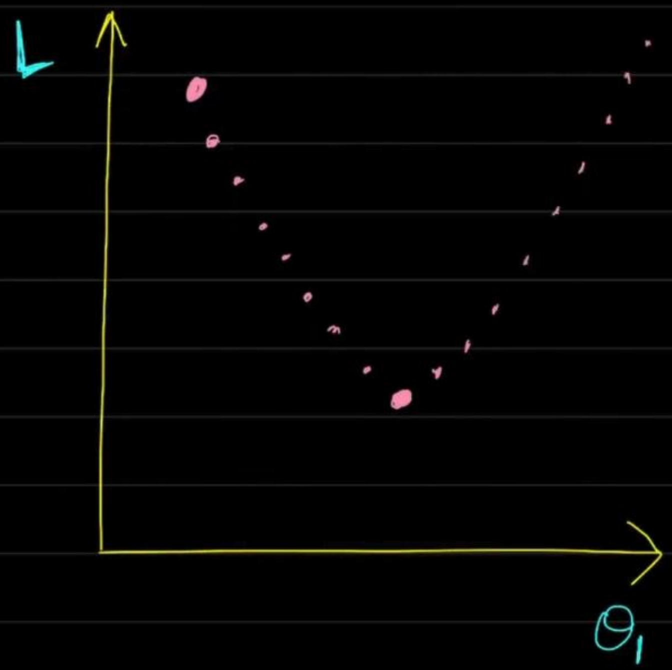
Curvature of the Loss function



Assume $\theta_0 = 0$

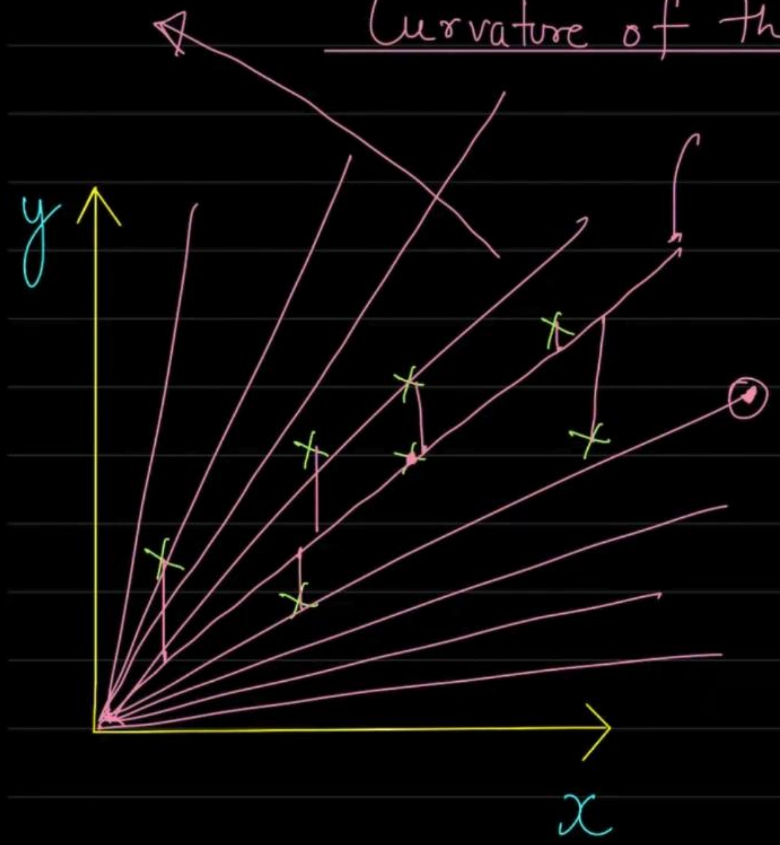
Curvature of the Loss function

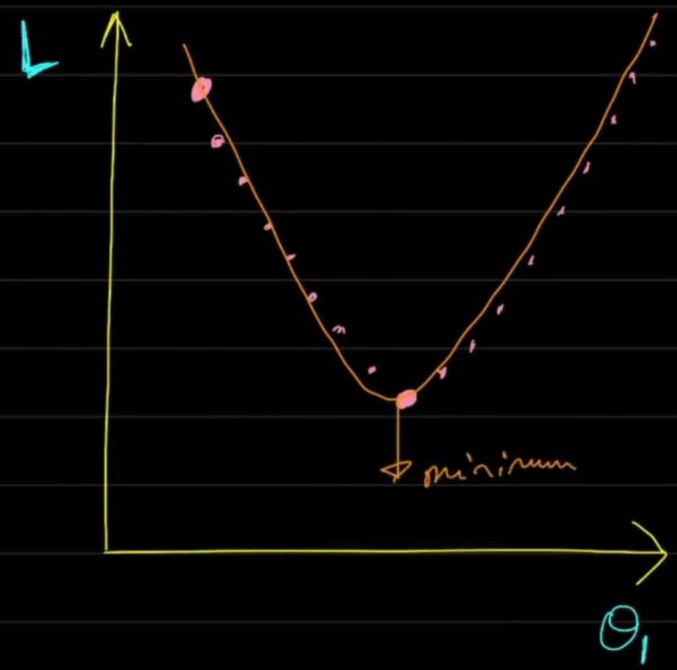




Assume $\theta_0 = 0$

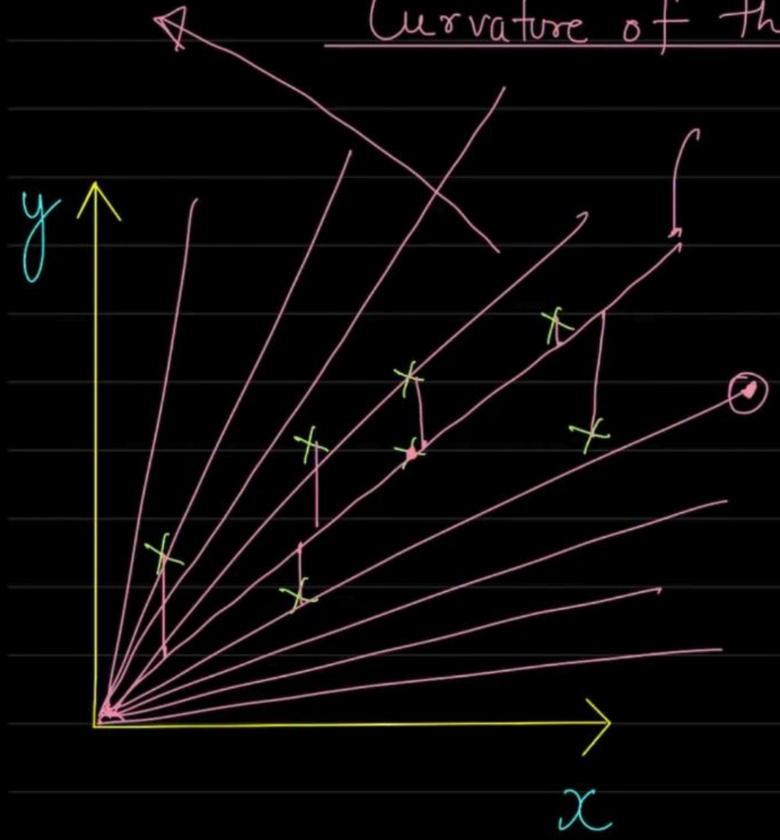
Curvature of the Loss function

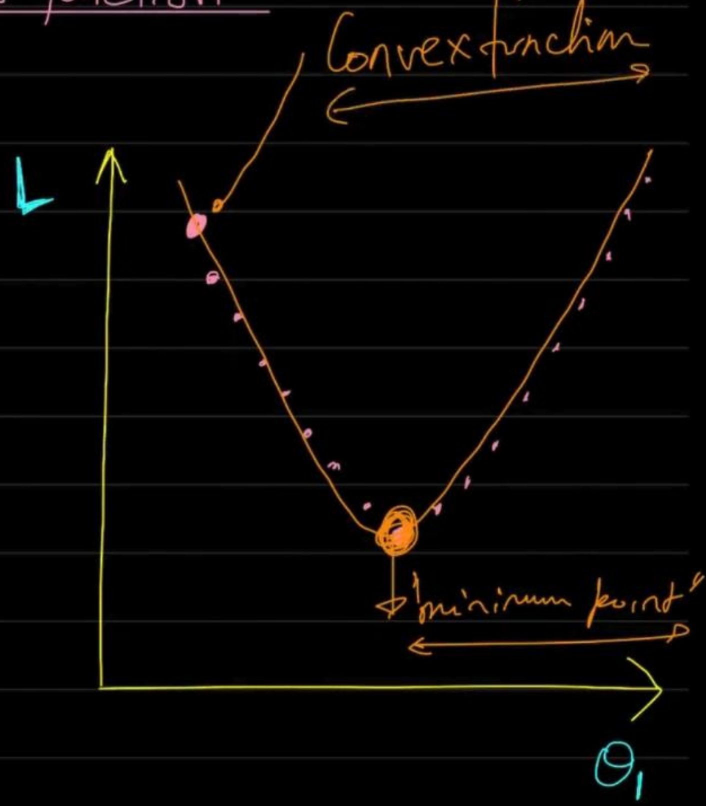




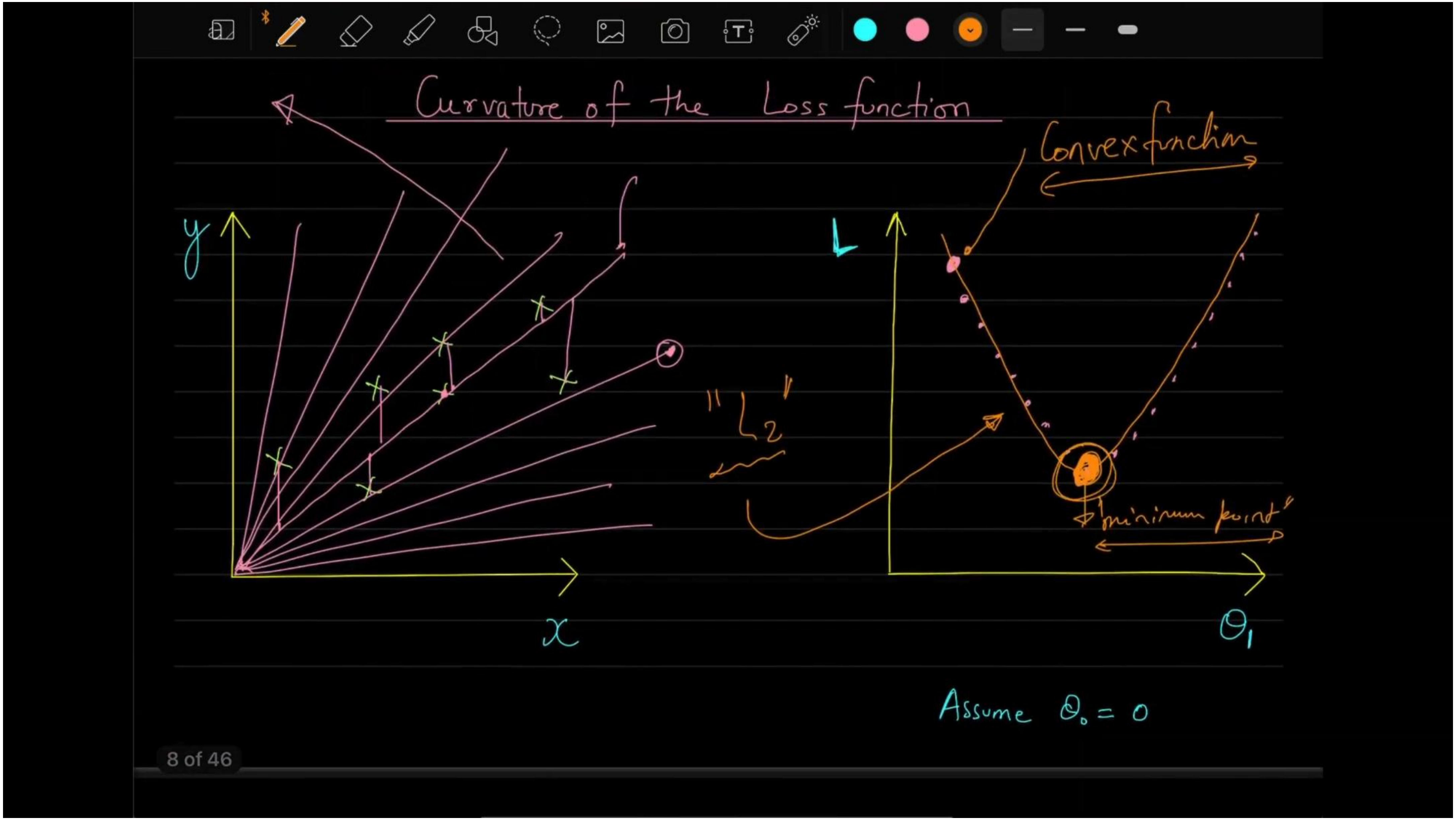
Assume $\theta_0 = 0$

Curvature of the Loss function

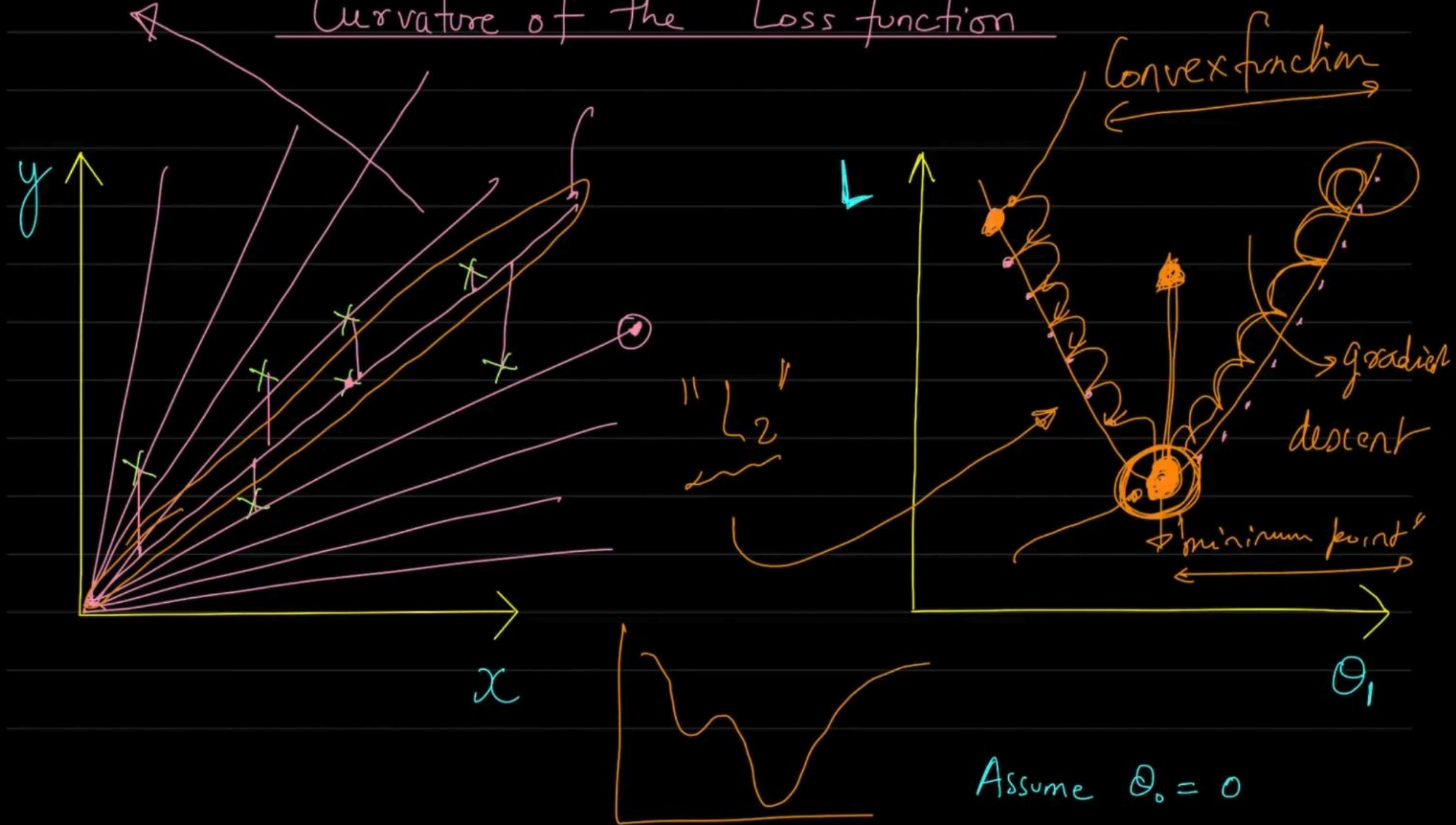




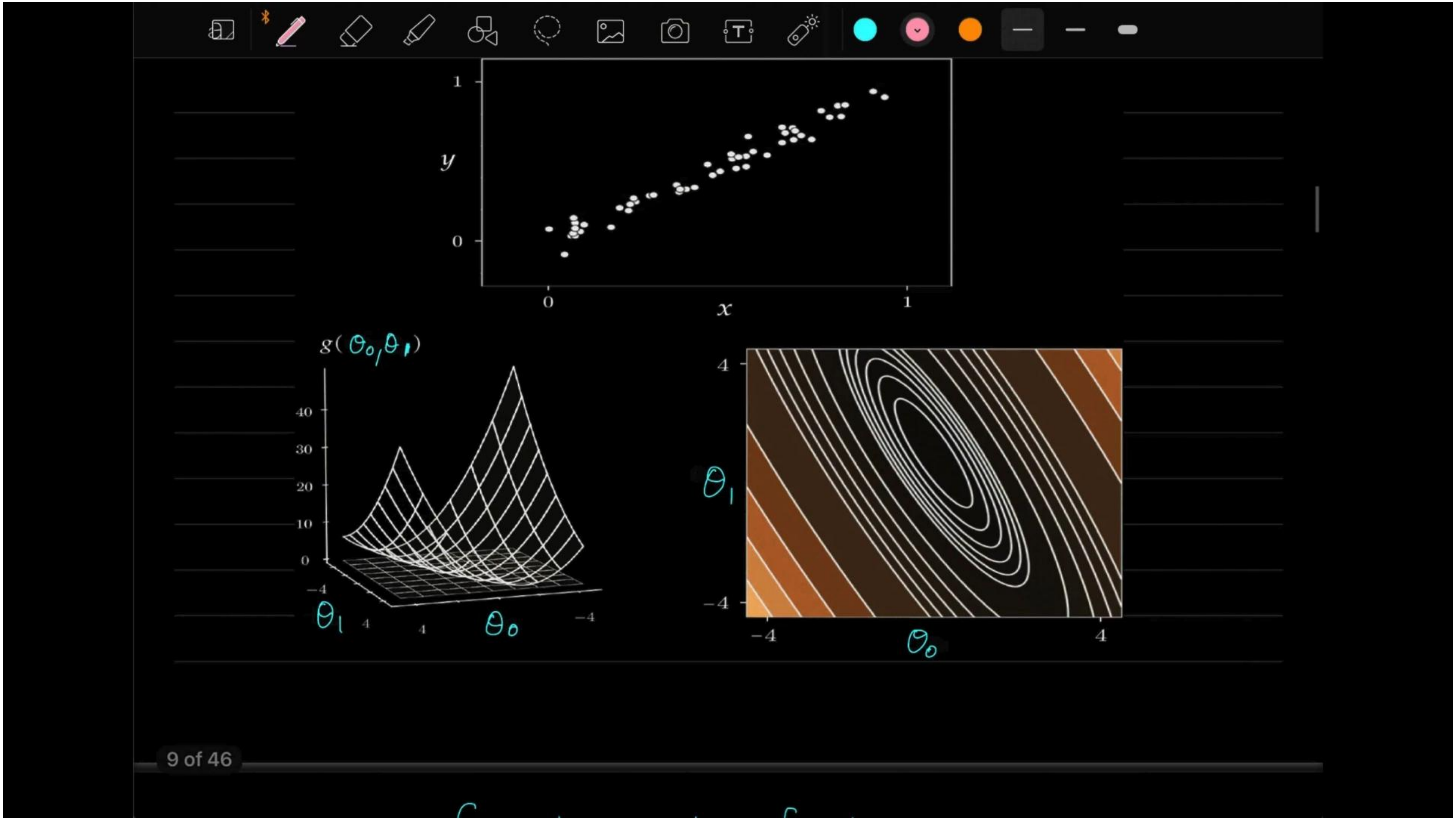
Assume $\theta_0 = 0$


Curvature of the Loss function
y
x
Convex function
L
"L2"
'minimum point'
θ1
Assume θ0 = 0
8 of 46

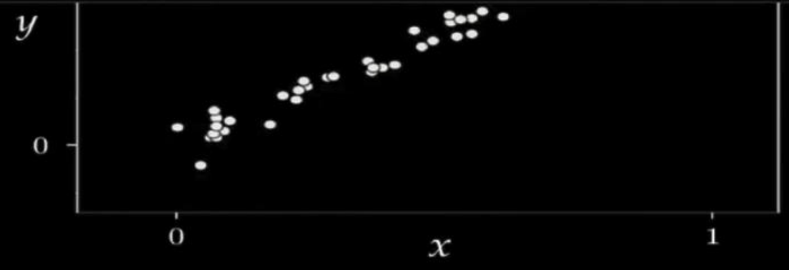
Curvature of the Loss function

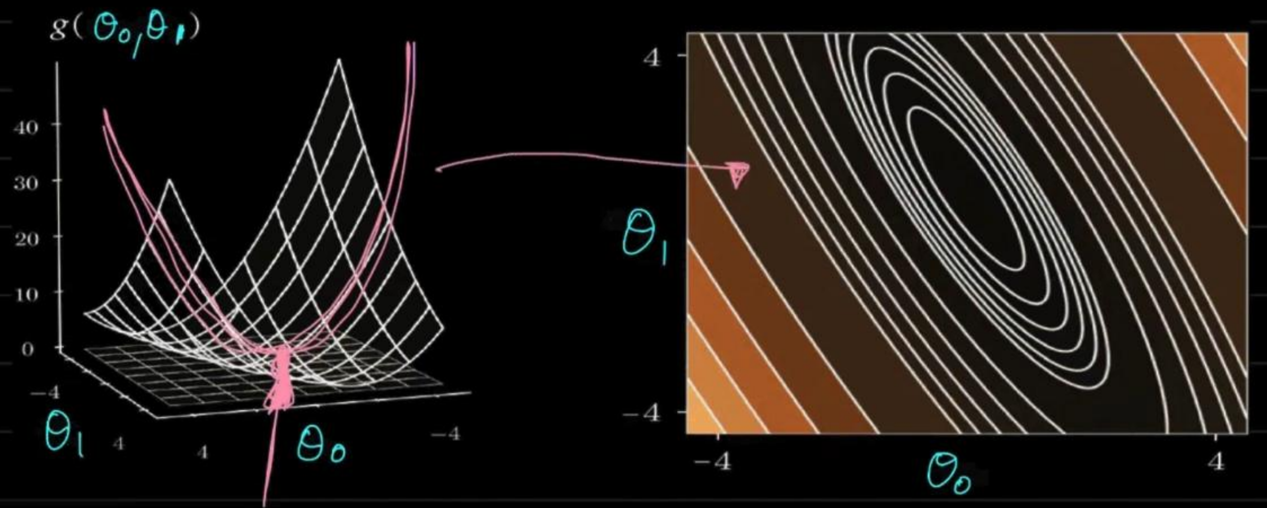


Assume $\theta_0 = 0$

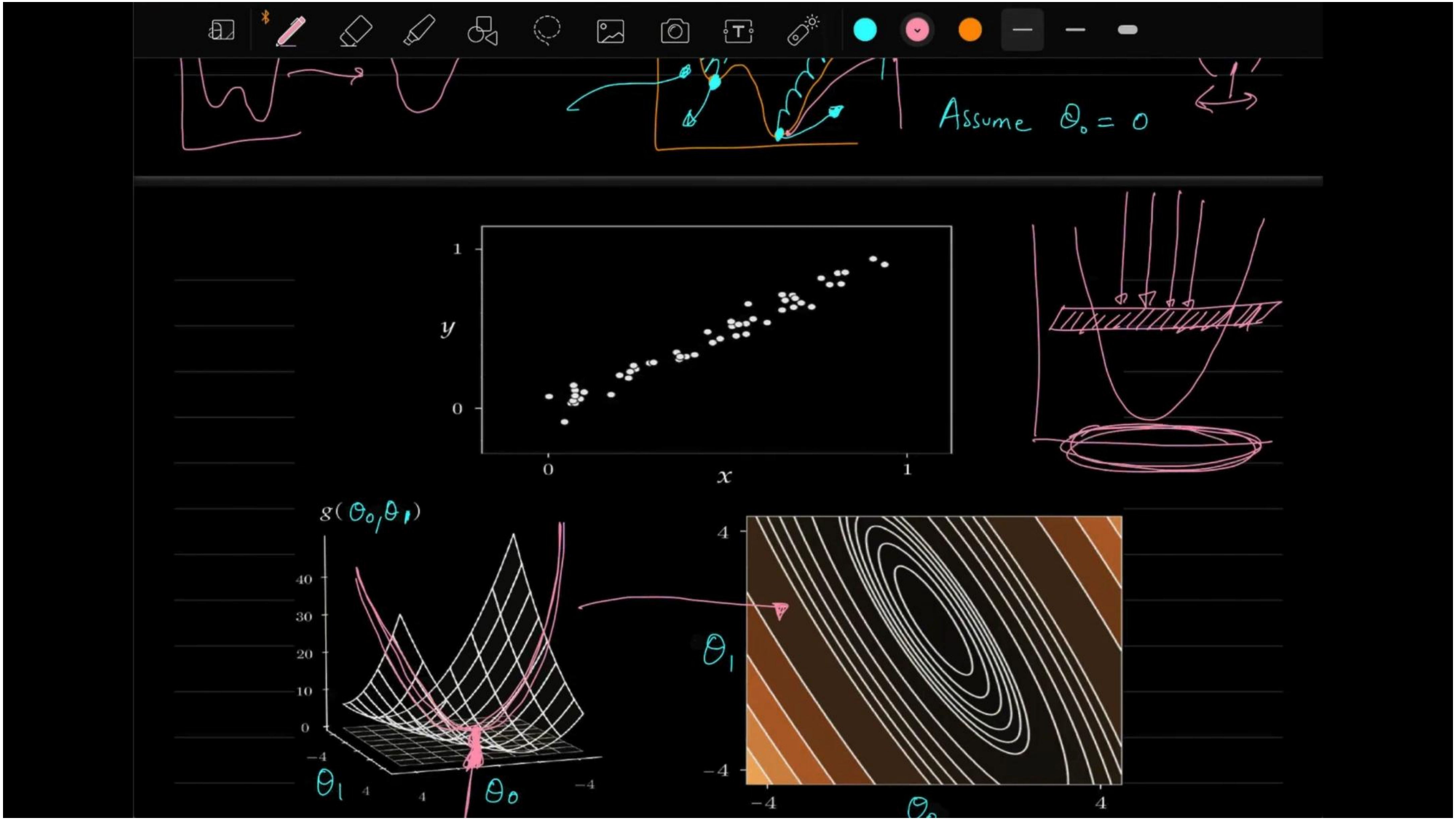


1
0
y
0
1
x
g(θ0,θ1)
40
30
20
10
0
−4
θ1
4
4
θ0
−4
4
−4
θ1
−4
θ0
4
9 of 46

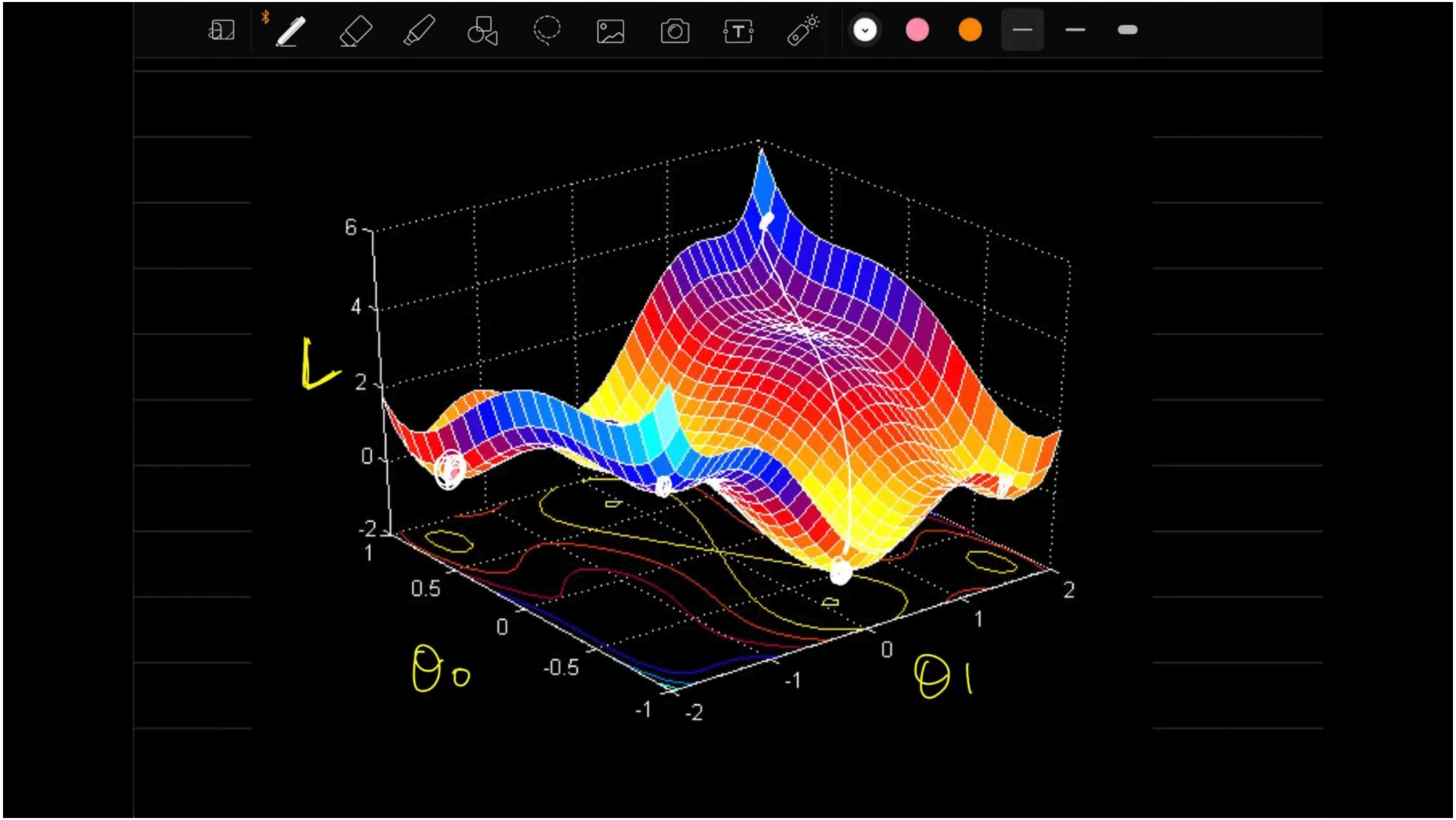




Generalized Loss function



Assume $\theta_0 = 0$
y
x
0
1
$g(\theta_0, \theta_1)$
40
30
20
10
θ_1
θ_0
-4
4



6
4
2
0
-2
L
1
0.5
0
-0.5
-1
θ0
-2
-1
0
1
2
θ1

* Two ways of obtaining optimal values of $\underbrace{\{\theta_0 \,\&\, \theta_1\}}$

(a) Numerical Method (Normal method)

(b) Using gradient descent

A Normal Method

$$L = \left\{ \frac{1}{2m} \sum_{i=1}^{m} \left(M_\theta(x^{(i)}) - y^{(i)} \right)^2 \right\}$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)} \right)^2 \qquad \text{—} \quad A$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \left[\left(\theta_0 + \theta_1 x_1^{(i)}\right)^2 + \left(y^{(i)}\right)^2 - 2\left(\theta_0 + \theta_1 x_1^{(i)}\right) y^{(i)} \right]$$

$$L = \frac{1}{2m} \sum_{i=1}^{m} \left[\theta_0^2 + \theta_1^2 \left(x_1^{(i)}\right)^2 + 2\theta_0 \theta_1 x_1^{(i)} + \left(y^{(i)}\right)^2 - 2\theta_0 y^{(i)} - 2\theta_1 x_1^{(i)} y^{(i)} \right]$$

$$\frac{\partial L}{\partial \theta_0}, \frac{\partial L}{\partial \theta_1} = ?$$

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{2m} \sum_{i=1}^{m} \left[2\theta_0 + 2\theta_1 x_1^{(i)} - 2 y^{(i)} \right]$$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{2m} \sum_{i=1}^{m} \left[2\theta_1 \left(x_1^{(i)}\right)^2 + 2\theta_0 x_1^{(i)} - 2 x_1^{(i)} y^{(i)} \right]$$

Equating the derivative to zero in order to obtain values of θ_0 & θ_1 that corresponds to the minimum of the loss function L

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{2m} \sum_{i=1}^{m} \left[2\theta_0 + 2\theta_1 x_1^{(i)} - 2y^{(i)} \right] = 0 \Bigg\}$$

$$\theta_0 + \frac{2}{2m} \sum_{i=1}^{m} \theta_1 x_1^{(i)} - \frac{1}{m} \sum_{i=1}^{m} y^{(i)} = 0$$

$$\theta_0 = \frac{1}{m} \sum_{i=1}^{m} y^{(i)} - \frac{\theta_1}{m} \sum_{i=1}^{m} x_1^{(i)}$$

$$\hat{\theta}_0 = \bar{y} - \theta_1 \bar{x}_1$$

To ensure minima we have to check the second derivative

To ensure minima we have to check the second derivative

$$\left(\frac{\partial^2 L}{\partial \theta_0^2}\right) = (2)\ (>0) \qquad (\because \text{ Minima is guranteed})$$

optimal value of $\hat{\theta}_1 = ?$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{2m} \sum_{i=1}^{m} \left[2\theta_1 \left(x_1^{(i)}\right)^2 + 2\theta_0\, x_1^{(i)} - 2\, x_1^{(i)} y^{(i)} \right] = 0$$

$$\theta_1 \frac{1}{m} \sum_{i=1}^{m} \left(x^{(i)}\right)^2 + \frac{1}{m} \sum_{i=1}^{m} \left(\bar{y} - \theta_1 \bar{x}_1\right) x_1^{(i)} - \frac{1}{m} \sum_{i=1}^{m} x_1^{(i)} y^{(i)} = 0$$

$$\theta_1 \frac{1}{m}\sum_{i=1}^{M}(x^{(i)})^2 + (\bar{y}-\theta_1\bar{x}_1)\frac{1}{m}\sum_{i=1}^{M}x_1^{(i)} - \frac{1}{m}\sum_{i=1}^{M}x_1^{(i)}y^{(i)} = 0$$

$$\theta_1\left[\frac{1}{m}\sum_{i=1}^{M}(x^{(i)})^2 - \bar{x}_1\bar{x}_1\right] = \frac{1}{m}\sum_{i=1}^{m}x_1^{(i)}y^{(i)} - \bar{y}\bar{x}_1$$

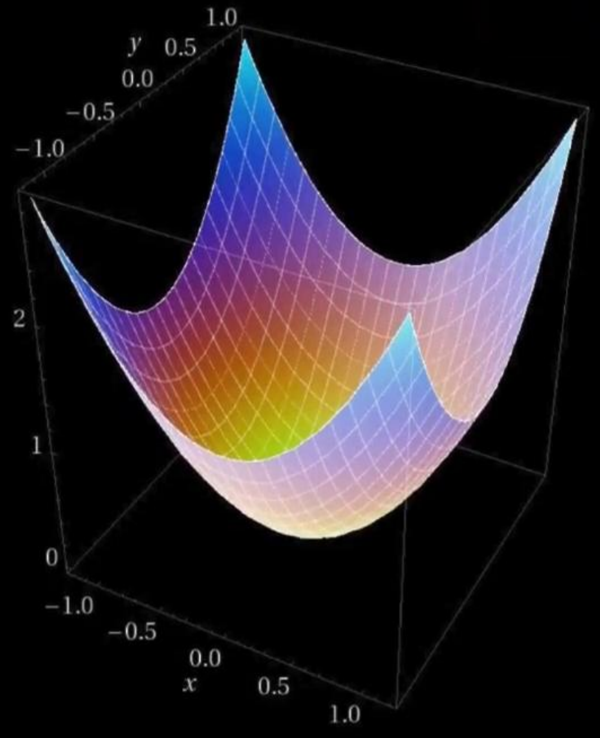
$$\hat{\theta}_1 = \frac{\frac{1}{m}\sum_{i=1}^{M}x_1^{(i)}y^{(i)} - \bar{y}\bar{x}}{\frac{1}{m}\sum_{i=1}^{M}(x^{(i)})^2 - \bar{x}_1\bar{x}_1}$$

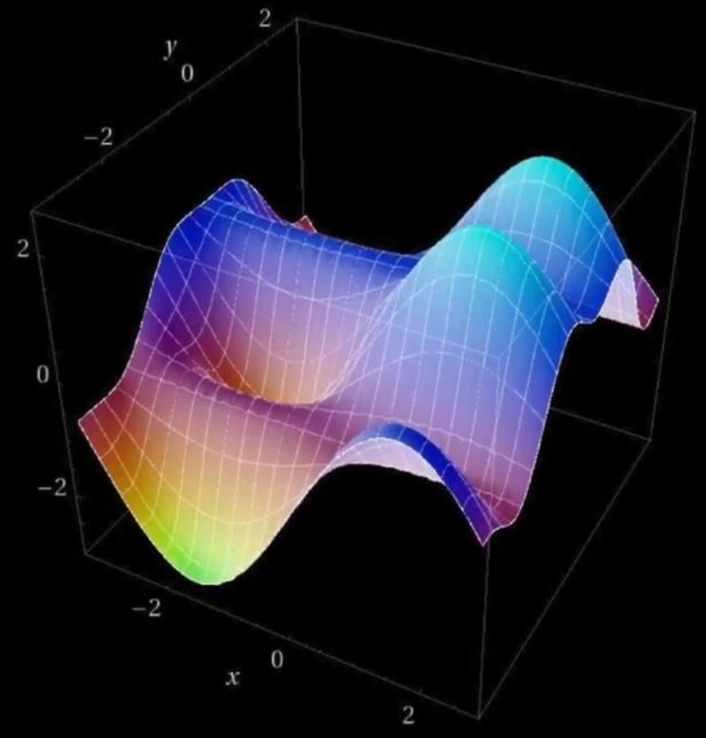
$$\hat{\theta}_1 = \frac{\frac{1}{m}\sum_{i=1}^{m} x_1^{(i)} y^{(i)} - \bar{y}\,\bar{x}}{\left(\frac{1}{m}\right)\sum_{i=1}^{m} \left(x^{(i)}\right)^2 - \bar{x}_1\,\bar{x}_1}$$

$$y = \hat{\theta}_0 + \hat{\theta}_1 x^{(i)}$$

Second Method: Gradient Descent

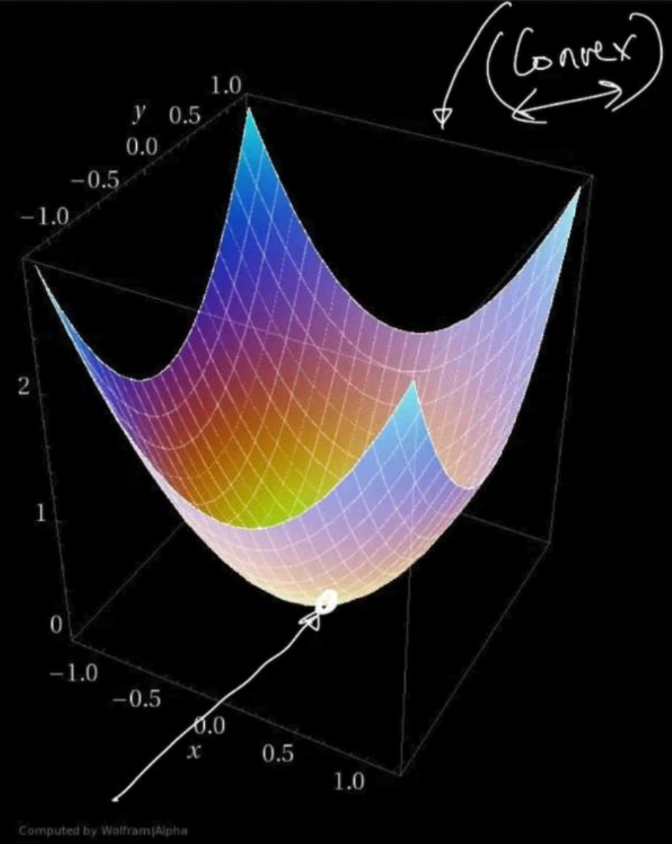
Second Method: Gradient Descent

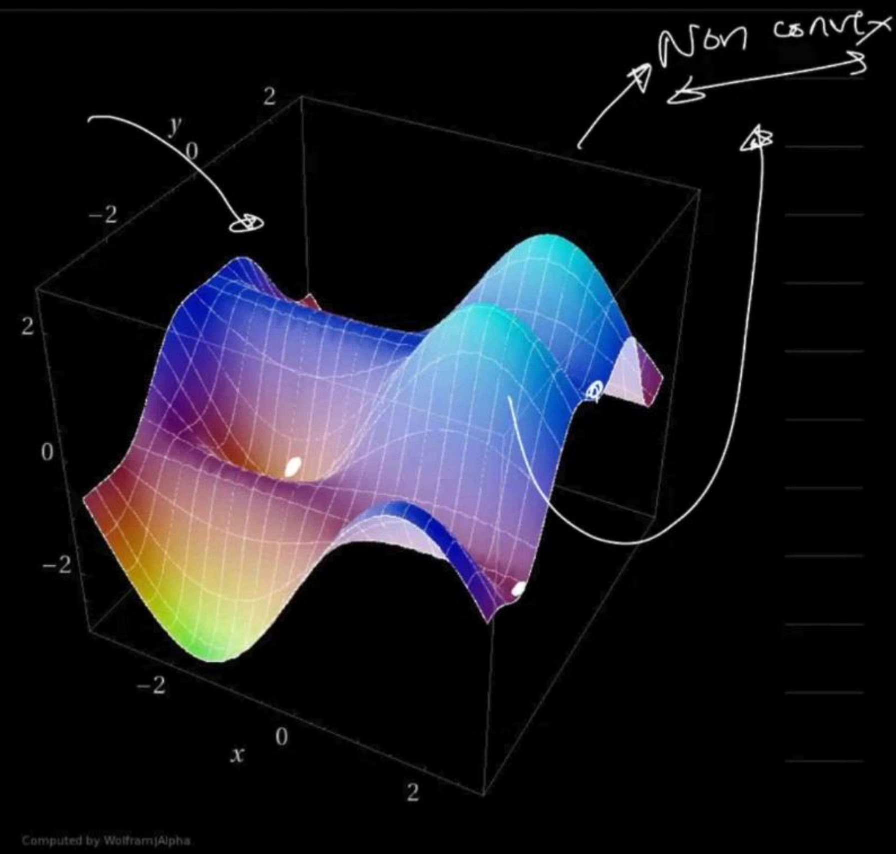


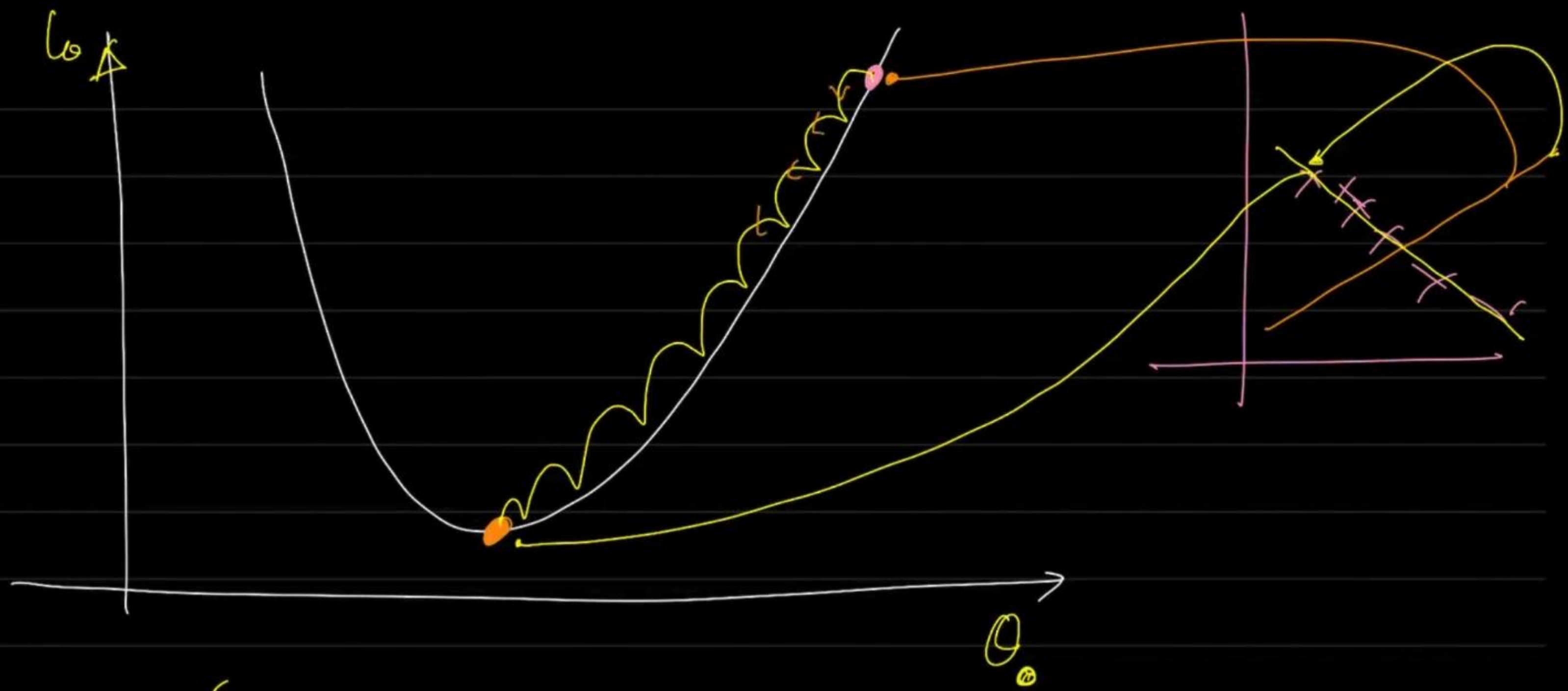


17 of 46

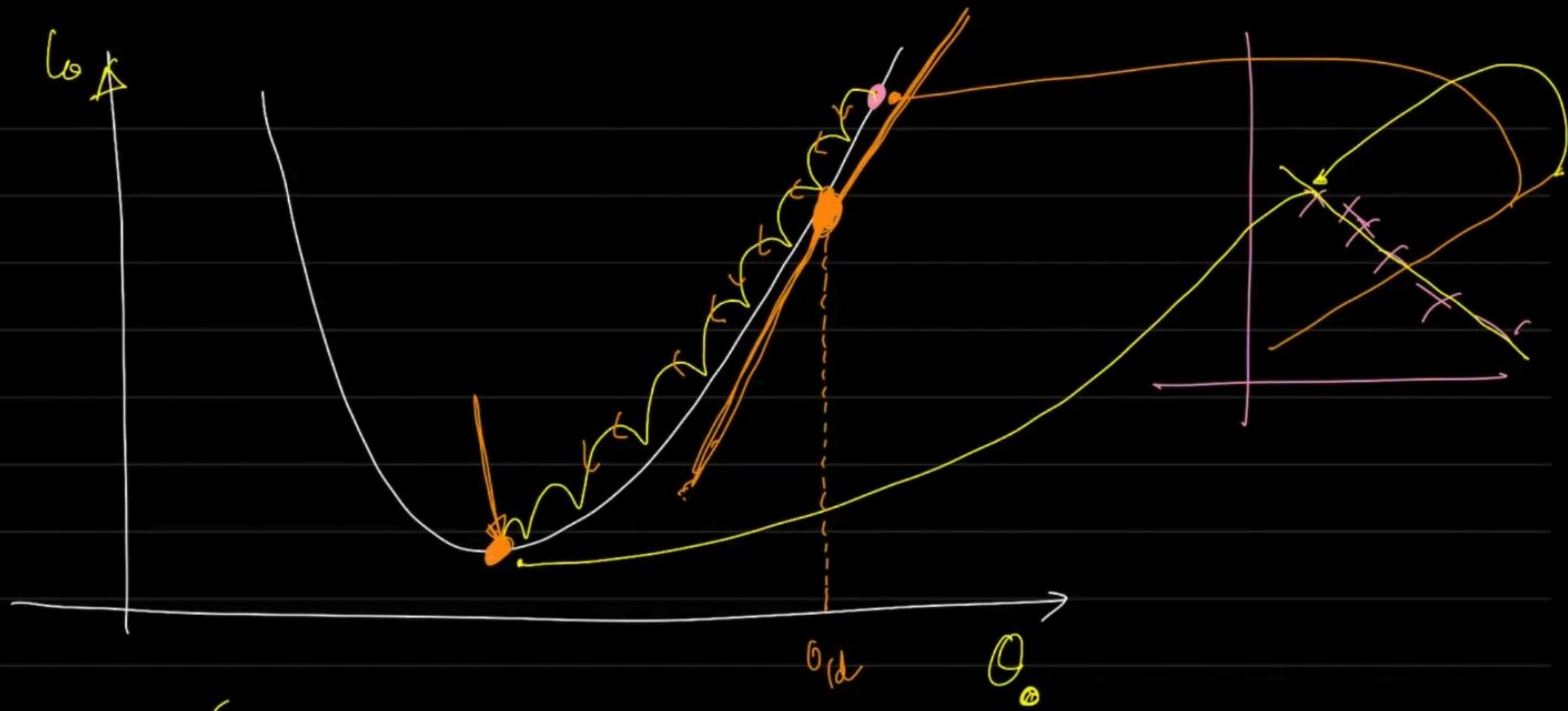
Second Method: Gradient Descent







$$\left\{ \theta_{new} = \theta_{old} - \alpha \frac{dL}{d\theta_{old}} \right\}$$



$$\theta_{new} = \theta_{old} - \alpha \frac{dL}{d\theta_{old}}$$