→ PAC < Finite Hypothesis / Infinite Hypothesis
→

\# Samples ✓

\# ϵ ✓

\# $(1-\delta)$ ✓

\# Complexity of the Hypothesis Class ✓

Finite Number of Samples in Training Data.

Data Augmentation

↳ Distribution

↳ Scaling, rotation, noise . . .

↳

→ Normal, Uniform, Poisson, ...

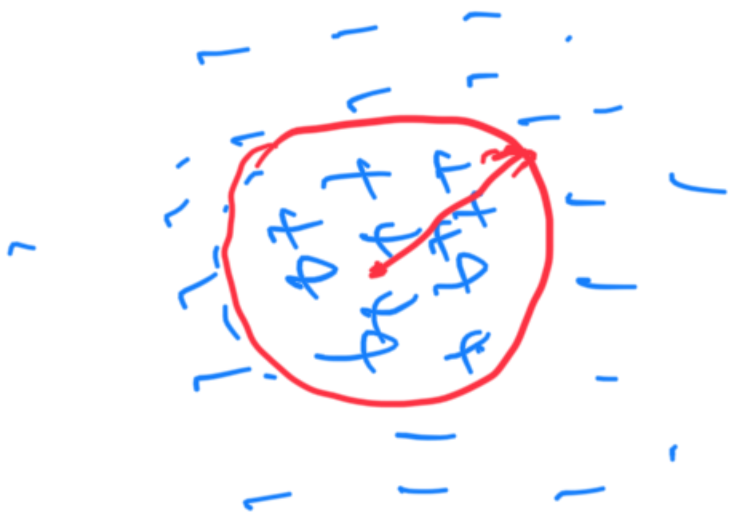
$x_i \in \mathbb{R}^d$

$y_i \in \mathbb{R}^1$

$$\begin{Bmatrix} x_i \\ y_i \end{Bmatrix} \xrightarrow{\pm \epsilon} \begin{Bmatrix} x_i' \\ y_i \end{Bmatrix}$$

$$x_i' = x_i \pm \epsilon$$

$\epsilon \in (0,1)^d$



1m + ( )

$\boxed{100}$ (−)

0.1m (−) m

x_1

x_1: We are happy

+E

we are hypppp

→ $\Sigma = \{a-z, A-Z, 0 \ldots 9, \ldots\}$

x_i: $\underbrace{100}$ character

→ $\underbrace{\text{Sentiment Classification}}$

joyful

Senti

x_1: We are happy. +ve

x_2: Mohan is playing. +ve

x_3: Chetan eat A sede in ML C +ve

x_4: She feels bad ✓ (-)ve ✓

Rohit is happy → $\boxed{m_1}$ ← +/-ve ✓

→ AI/ML/DL → Application

↳ ML AAS

↳ Development Tools

Linear Regression

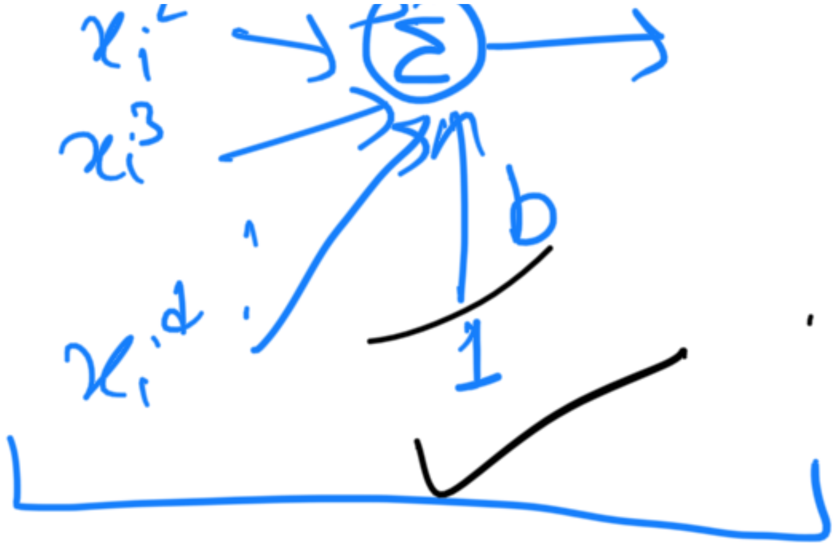
$x_i \in \mathbb{R}^d$; $y_i \in \mathbb{R}^1$

$h \in H$

$x_i \rightarrow \hat{y}_i = h_\theta(x_i)$

$h(x_i) = \theta^T x_i + b$

$$h_\theta(x_i) = \underbrace{\theta\, x_i + b}_{\theta^T x_i}$$



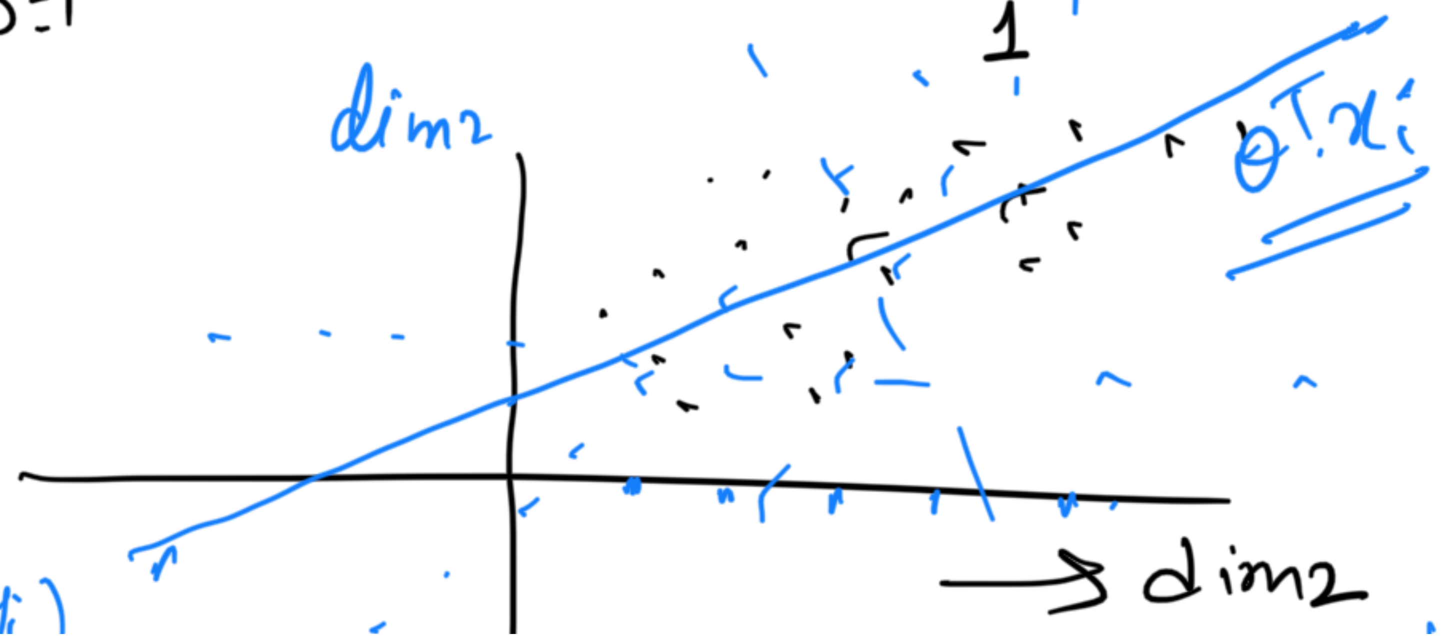
$$x_i = \begin{bmatrix} 1 \\ x_i^1 \\ x_i^2 \\ \vdots \\ x_i^d \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}$$

$$\theta^T x_i = \theta_0 \cdot 1 + \theta_1 \cdot x_i^1 + \theta_2 \cdot x_i^2 + \dots \theta_d \cdot x_i^d$$

$$= \sum_{j=1}^{d} \theta_j x_i^j + \underline{\theta_0}$$

$x_i \in \mathbb{R}^2,\ y_i \in \mathbb{R}^1$

$\{x_i, y_i\}_{i=1}^N$



| f1 | f2 | label(y_i) |

x_1

x_2

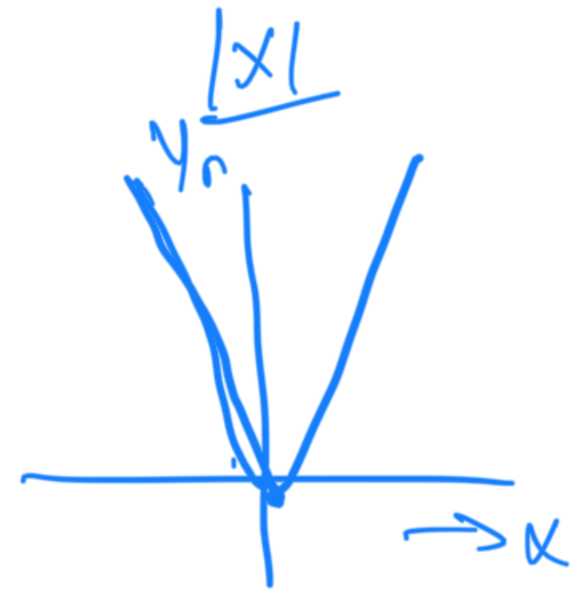
$\vdots$

x_N

True Label = $\underbrace{y_i}$

Predicted

$$\underbrace{\hat{y}_i} = \underbrace{h_\theta(x_i) = \theta^T \cdot x_i}$$



Absolute error (x_i) = (e_i) = $|y_i - \hat{y}_i|$

Mean total error$(h) = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$

squared error $(e_i) = (y_i - \hat{y}_i)^2$

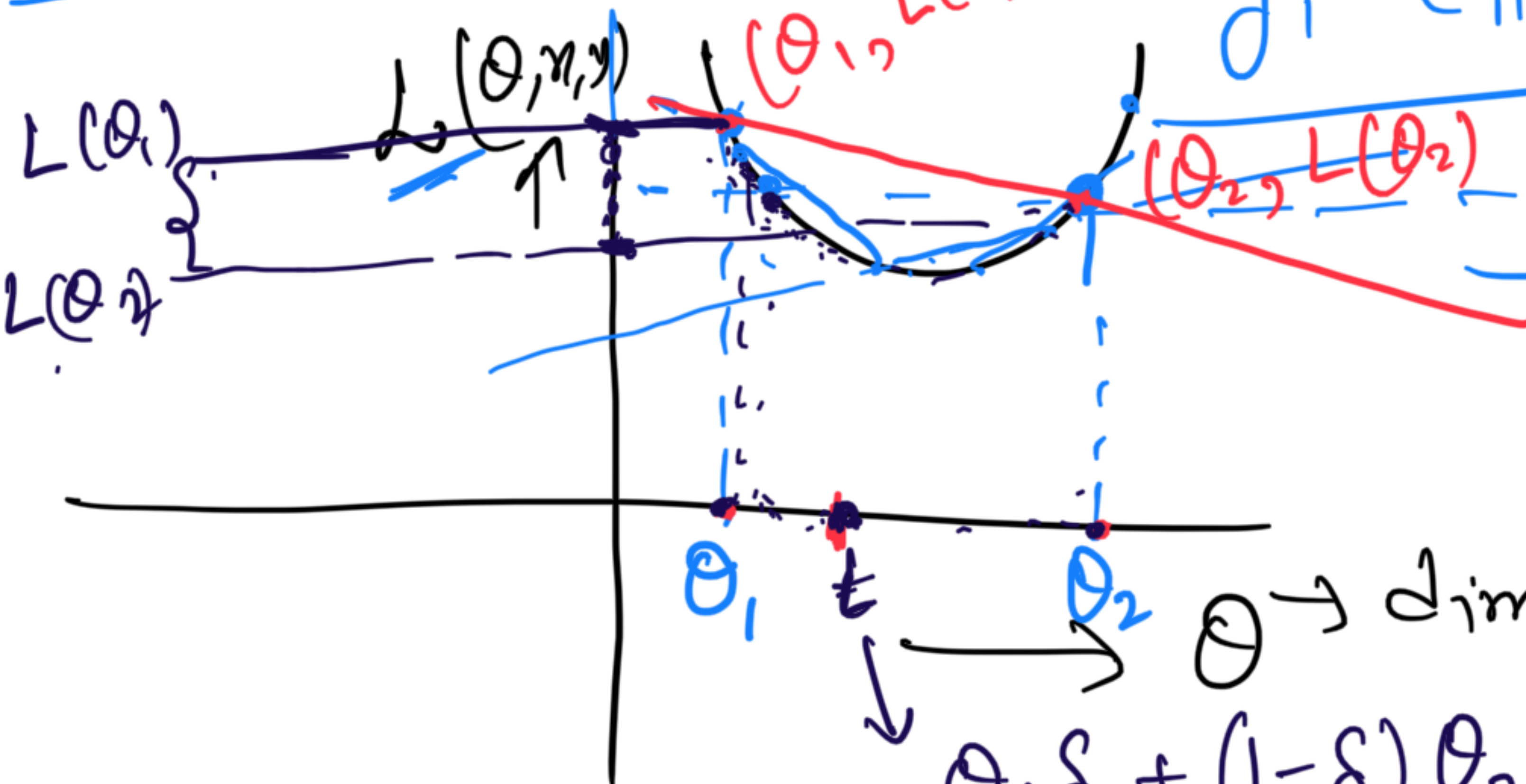
Mean Squared Error (mse) $= \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

loss function $= \frac{1}{N} \sum_{i=1}^{N} (y_i - \theta^T x_i)^2$

$h(\theta, x, y) = \underbrace{\qquad\qquad}$

$y_i \in \mathbb{R}^1$

$-\infty, +\infty$



$\delta \in [0,1]$

$\theta \rightarrow$ dim

$\underbrace{\theta_1 \delta + (1-\delta)\theta_2}$

$L($

$$\boxed{\theta = (X^T \cdot X)^{-1} \cdot X^T y}$$

$$L(\theta_1 \delta + (1-\delta)\theta_2) \leq \delta L(\theta_1) + (1-\delta) L(\theta_2)$$

$$\underline{y = x^2}$$

$$\frac{dy}{dx} = 2x = 0$$

$$\Rightarrow \underbrace{x = 0}$$

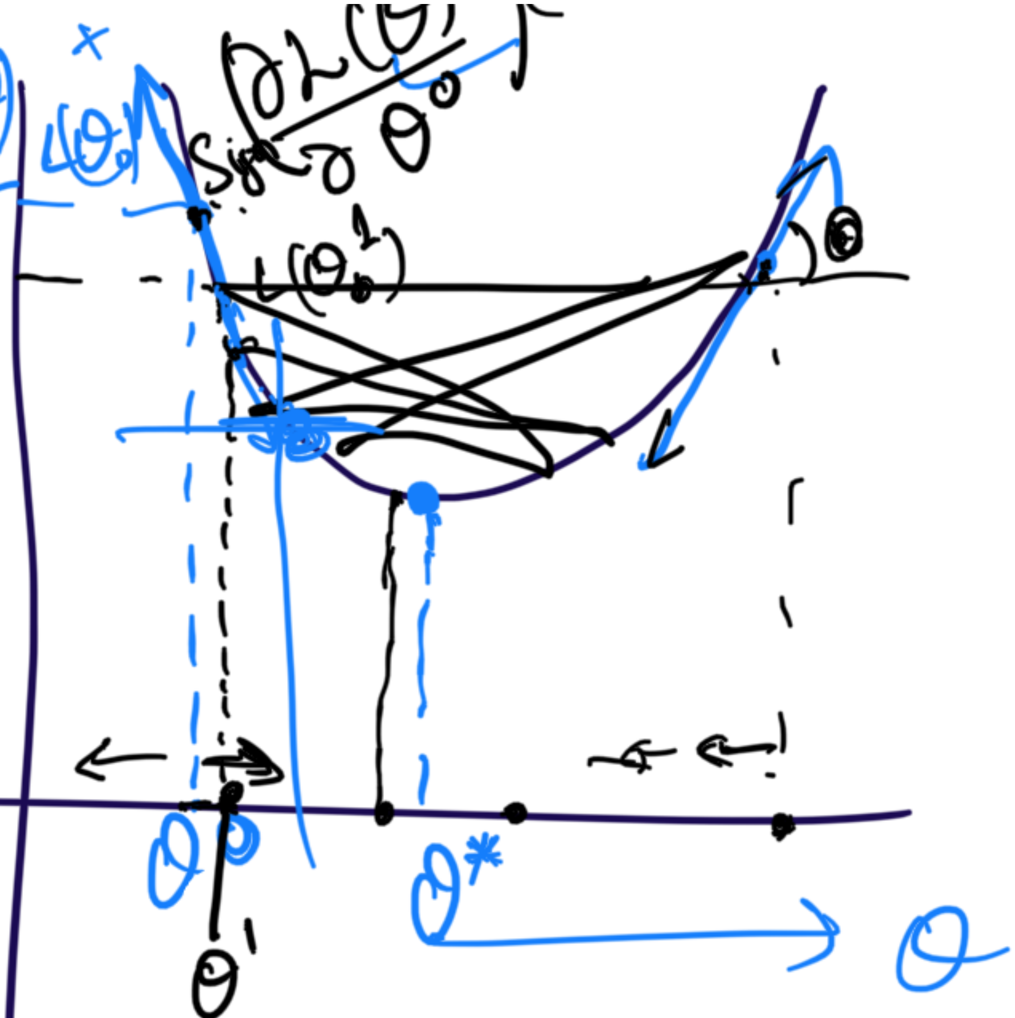
$$\frac{d^2y}{dx^2} = 2 \quad +ve \qquad \underline{\text{minima}}$$

$$y = (x^1)^2 + (x^2)^2$$

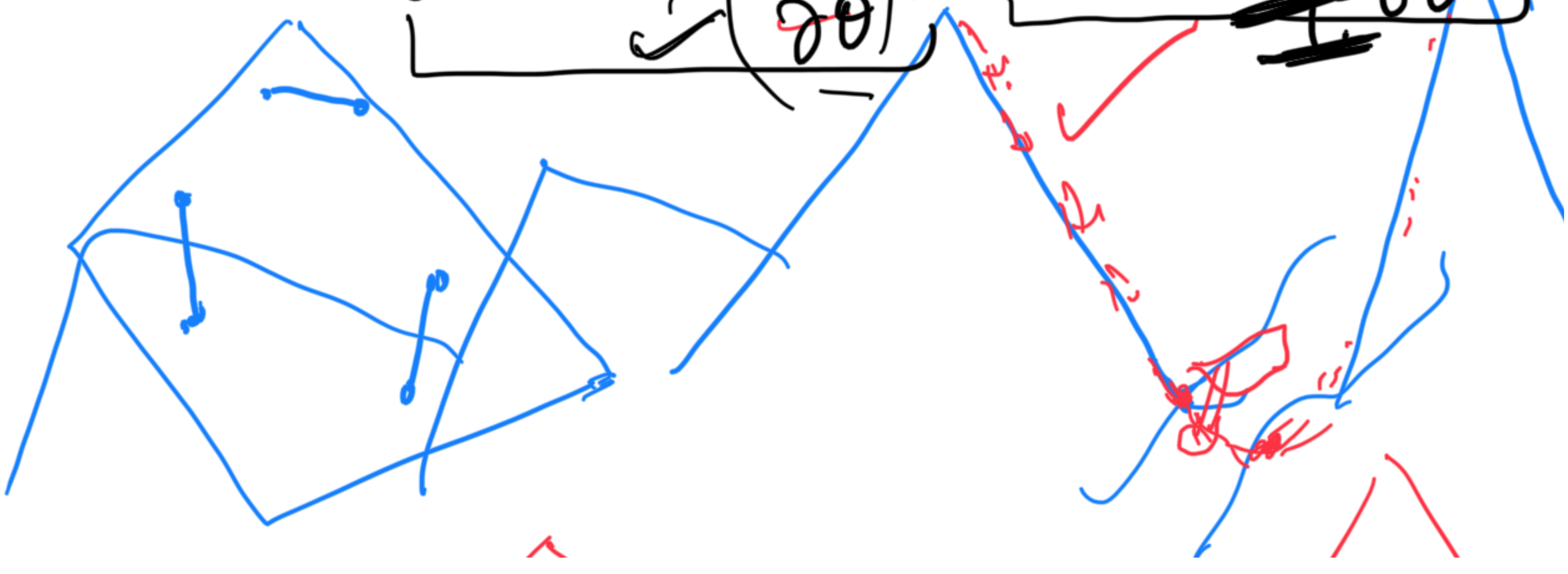
$$\frac{\partial y}{\partial x^1} = 2.x^1$$

$$\frac{\partial y}{\partial x^2} = 2.x^2$$

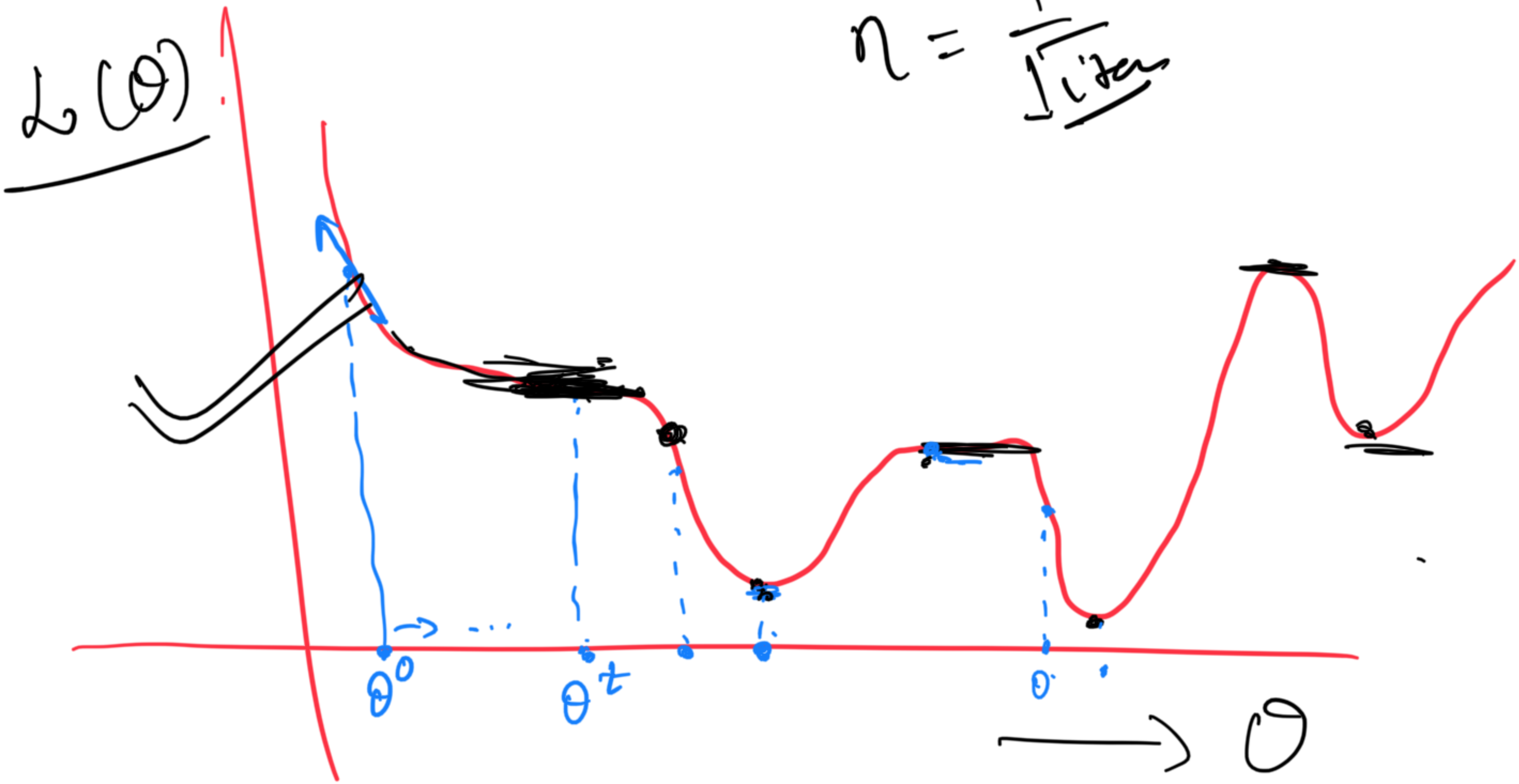
$c_1 \; |x| + c_2$



$$\underbrace{\theta^1 = \theta^0 - \left(\frac{\partial L}{\partial \theta}\right)} \qquad \underbrace{\theta^1 = \theta^0 - \eta \frac{\partial L(\theta)}{\partial \theta^0}}$$



$$\eta = \frac{1}{\sqrt{iter}}$$



$\vec{v}$

$$\frac{\vec{V}}{\|\vec{V}\|}$$

$$L(\theta) = \theta^2$$

$$\theta^0$$

$$\theta^1 = \theta^0 - \underbrace{\frac{\partial L(\theta)}{\partial \theta^0}}$$

$$= \theta^0 - 2.\theta^0$$

$$= -\theta^0$$

$$\theta^t = \theta^{t-1} - \frac{\partial L(\theta)}{\partial \theta^{t-1}}$$

$$\theta^t = \underbrace{\theta^{t-1} - \eta . \frac{\partial L(\theta)}{\partial \theta^{t-1}}}$$

$$\theta^2 = \underline{\theta}^1 - 2\theta^1 = -\theta^1$$

$$\eta = 0.01$$

for $t = 1$ to max_iteratio {

for each parameter θ_j {

$$\theta_j^{t} = \underbrace{\theta_j}_{} - \eta \underbrace{\frac{\partial L(\theta)}{\partial \theta_j}}_{}$$

end for

if $|L(\theta_j^{t}) - L(\theta_j^{t-1})| \leq 10^{-12}$ ||

$$\left| \frac{\partial L(\theta)}{\partial \theta^{t-1}} - \cancel{\frac{\partial L(\theta)}{\partial \theta^{t-1}}} \right| \leq 10^{-15}$$

exit / break.

end for

$$L(\theta) = \frac{1}{\ldots} \sum^{N} (\ldots \theta^T \ldots)^2$$

$$L(\theta) = \frac{1}{N} \sum_{i=1} \underbrace{(y_i - \theta x_i)}$$

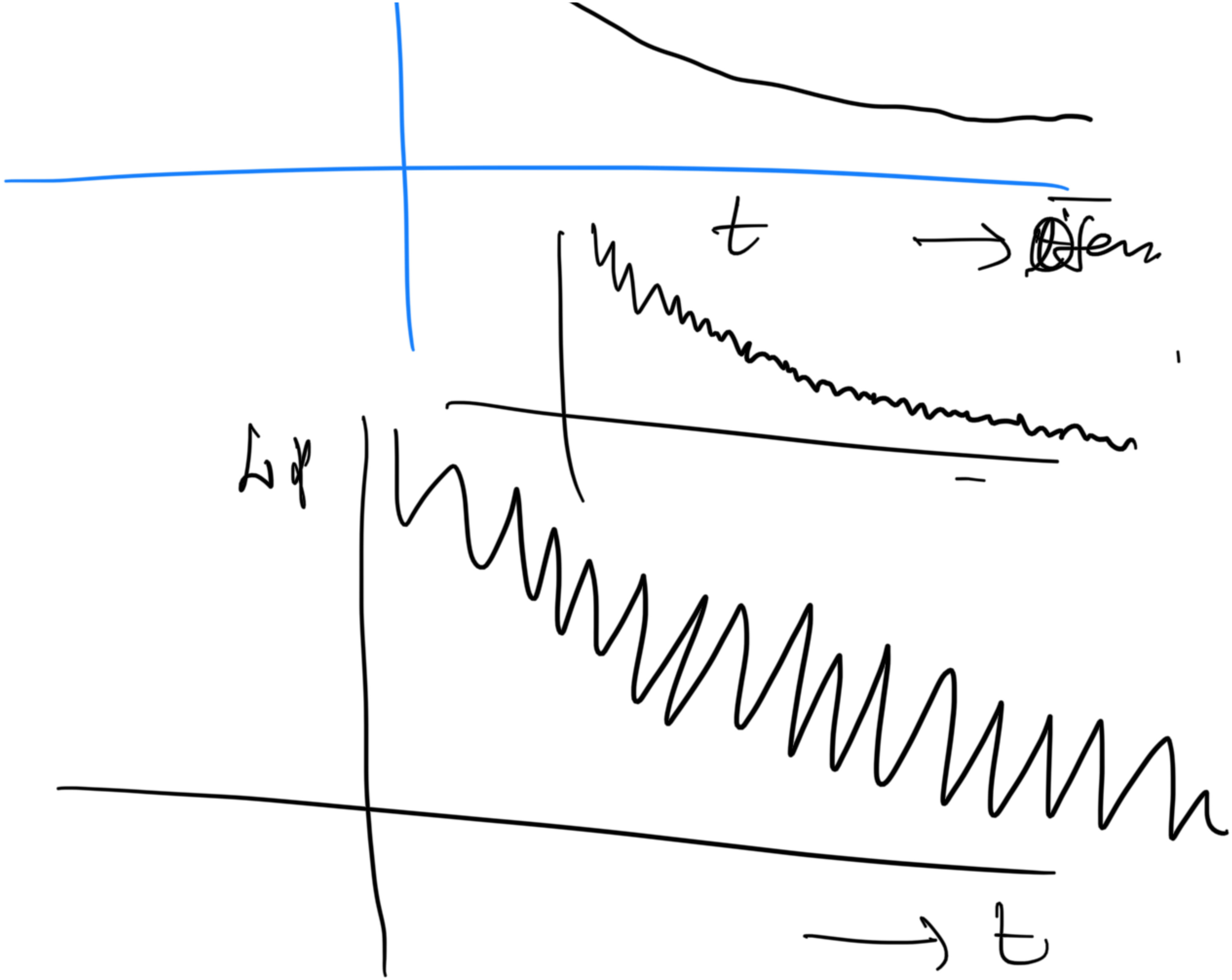
$$\frac{\partial L(\theta)}{\partial \theta_j} = \underbrace{-\frac{2}{N} \sum_{i=1}^{N} \underbrace{(y_i - \theta^T x_i) \cdot x_i^j}}$$

Batch_size

$$O(Nd)$$

$$\theta_j^t = \theta_j^{t-1} - \eta \cdot \frac{\partial L(\theta)}{\partial \theta_j}$$

$L(\theta)$


t
t

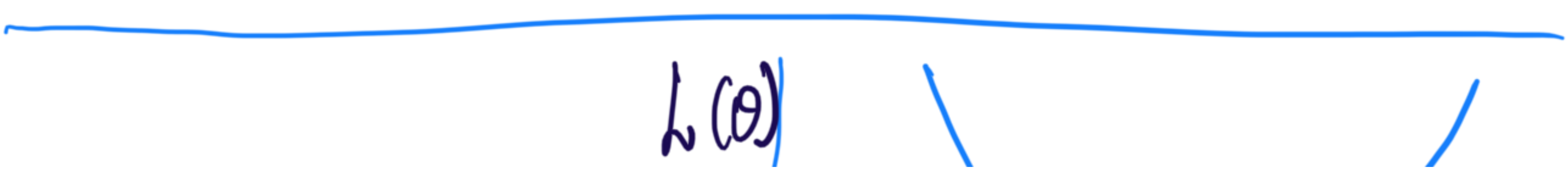
$$\delta = -\eta \frac{\partial L(\theta)}{\partial \theta^{0}} \qquad \gamma \in \underline{(0,1)}$$

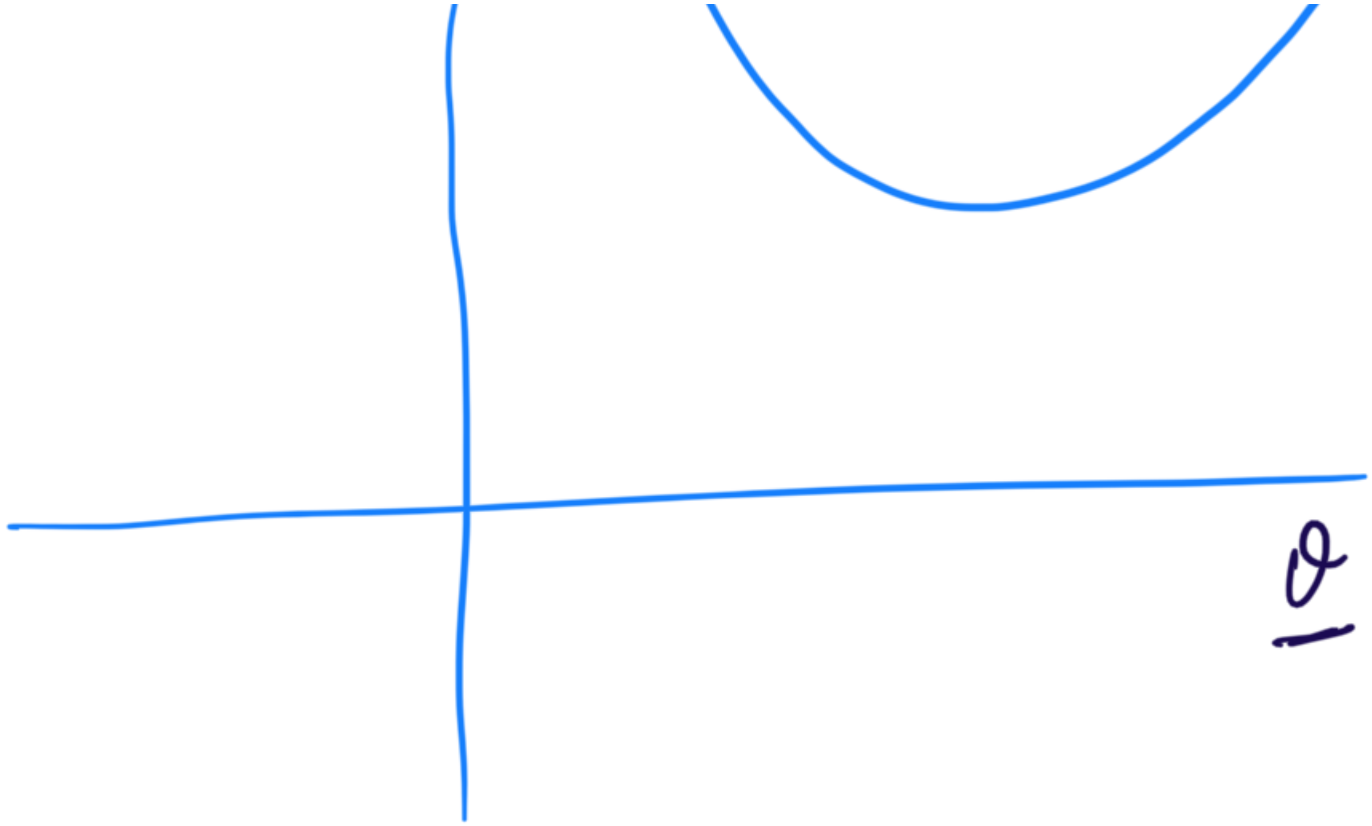
$$\delta^{t} = \gamma \cdot \delta^{t-1} + \eta \frac{\partial L(\theta)}{\partial \theta}$$

momentum, Adagrad, RmsProp

<u>Adam</u>

$$\frac{\partial L(\theta)}{\partial \theta_j} \qquad \underset{\theta}{\text{minimize}} \ \frac{1}{N}\sum_{i=1}^{N}(y_i - \theta^T x_i)^2$$





$$\underline{\mathcal{L}(\theta, x_i, y_i)}$$

$$\left[\max_{\|\delta\| < \epsilon} \mathcal{L}(\theta, x_i + \delta, y_i)\right.$$

$$\epsilon > 0$$

$$x^i_{adv} = x_i + \delta^*.$$

$$x_{adv} = \underbrace{x_i + \epsilon \cdot \frac{\partial L(\theta)}{\partial x_i}}$$