Doubt - Class-2

2 degree

Polynomial

$$\Rightarrow f(x) \simeq f(a) + g^T (x-a) + \frac{1}{2}(x-a)^T H (x-a)$$

Doubt: Why H is Symmetric ?

$$g = \nabla_x f(x) \qquad H = \nabla_x^2 f(x)$$

↳ Symmetric

x - vector

$[x_1 \; x_2 \; \dots \; x_n]$ $\xrightarrow{\nabla_x f(x)}$ $f(x) \Rightarrow$ ~~Vector~~ Scalar fn.

$$f(x) = (1 \times 1)$$

$$[\eta_1 \dots \eta_n] \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \dots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \dots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$D \times 1$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$D \times 1$

$D \times 1 \times (1 \times D)$

$= D \times D$

$1 \times D$
$D \times 1$

$$\begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_1}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial x_1}\right) & \dots & \frac{\partial}{\partial x_D}\left(\frac{\partial f(x)}{\partial x_1}\right) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_2}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial x_2}\right) & & \frac{\partial}{\partial x_D}\left(\frac{\partial f(x)}{\partial x_2}\right) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_3}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial x_3}\right) & & \frac{\partial}{\partial x_D}\left(\frac{\partial f(x)}{\partial x_3}\right) \\ \vdots & & & \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f(x)}{\partial x_D}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f(x)}{\partial x_D}\right) & \dots & \frac{\partial}{\partial x_D}\left(\frac{\partial f(x)}{\partial x_D}\right) \end{bmatrix}$$

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial f(x)}{\partial x_i}\right)$$

$$\begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_D \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_D \partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_D} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_D} & \cdots & \frac{\partial^2 f(x)}{\partial x_D^2} \end{bmatrix}$$

$\nabla^2 f(x)$ — Hessian

$$y^{(i)} = \theta^T x^{(i)} + \epsilon \quad \rightarrow \epsilon \sim N(0, \sigma^2)$$

doubt

↳ why normal dist?

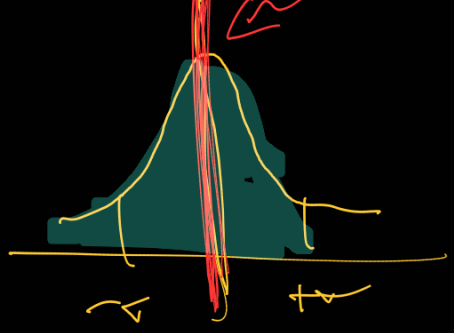
[When we do not have any idea of underlying distribution over some variable, we take it as Normal dist^n because N.D is "universal" in nature and any phenomena can be explained by using N.D as a proxy.]

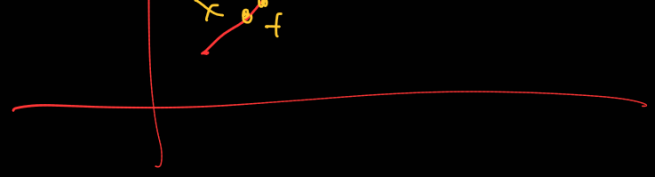
→ Why $\mu = 0$?

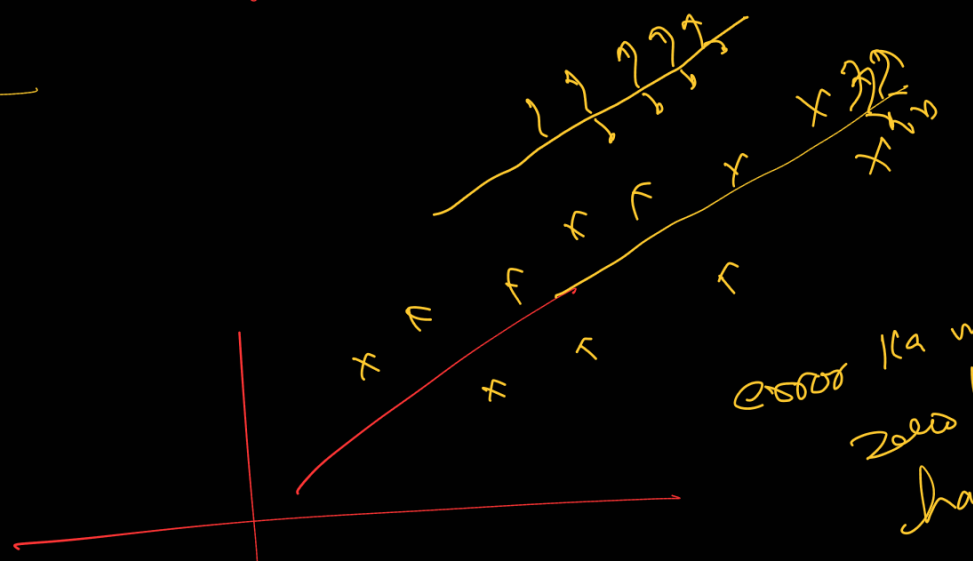
$$y^{(i)} = \theta^T x^{(i)} + \epsilon$$

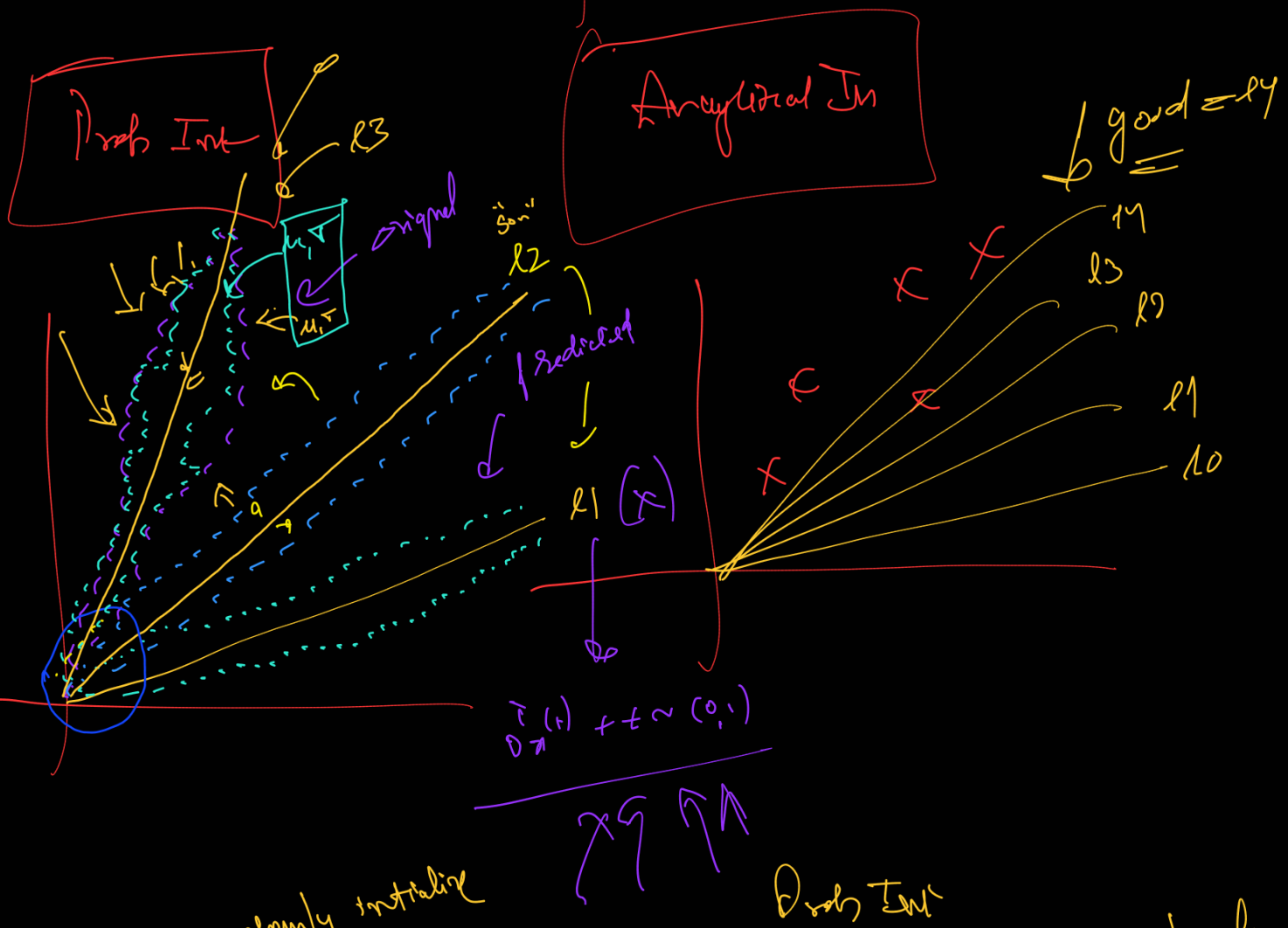
$\epsilon \sim N(0, 1)$

good distribution









θ = randomly initialize

↳ 'θ_{MLE}'

"θ_{MLE}"

Prob Int

↳ I am trying to learn

eq of actual data distribution

हम same ही kaam kar rahaa hu

I am reducing the same loss.

मेरा जो estimate है data का, उसका likelihood maximize होता है।

↳ "θ_{MLE}"

doubt: L का eq^n कहाँ से आया।

$$L = \phi^{y} (1-\phi)^{(1-y)}$$ → Bernoulli distribution

Binomial Distn → Repeated Bernoulli trials

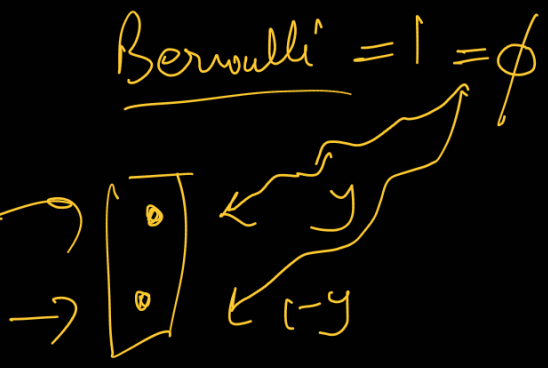
$$^{n}C_{r}\, p^{r} (1-p)^{n-r} \quad = n$$

p ↓ ϕ

Total it should occur

Bernoulli = 1 = ϕ

$\phi^{y} (1-\phi)^{(1-y)}$



$$\theta_{MLE} = \arg\max_{\theta} \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}; \theta)$$

$\phi^{y} (1-\phi)^{(1-y)}$

I am trying to maximize the likelihood of all the points inside my training data.

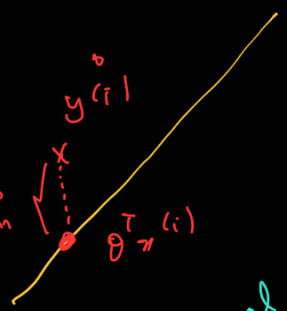
equivalent

Minimizing $J(\theta)$ ← Gradient Descent

5 datapoints

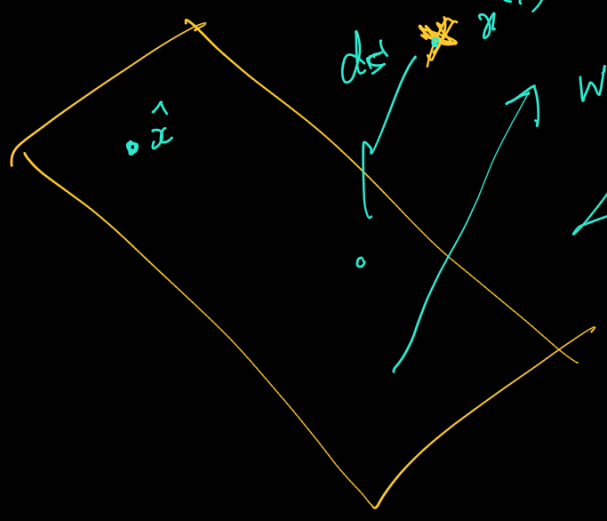
doubt: Why distance?

$(y^{(i)} - \theta^T x^{(i)})^2$



normal to the plane

$d^{(i)} = w^T (x^{(i)} - x')$



$$d = \vec{w}\,(\vec{x}^{\,p} - \vec{x})$$

(TEF)

Vector 3D $\rightarrow$ x11 (TEF)

$$d^{(i)} = w^T\left(x^{(i)} - x\right)$$

$x^{(i)}$ x

$\rightarrow$ I will teach you

$$J(\theta) = (y - X\theta)^T (y - X\theta)$$

$$= (y^T - \theta^T X^T)(y - X\theta)$$

$$= y^T(y - X\theta) - \theta^T X^T (y - X\theta)$$

$$= y^T y - y^T X\theta - \theta^T X^T y + \theta^T X^T X\theta$$

$\nabla_\theta \theta^T A\theta = 2A\theta$

$\nabla_\theta a^T\theta = a$

$\nabla_\theta (a^T\theta) = a^T$

$$y^T y - Y^T X\theta - \theta^T X^T Y + \theta^T X^T X\theta$$

$$y^T y - y^T X\theta - y^T X\theta + \theta^T X^T X\theta$$

$$= y^T y - 2y^T X\theta + \theta^T X^T X\theta$$

$$Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta$$

$$J(\theta) \quad Y^T Y - 2(X^T Y)^T\theta + \theta^T X^T X\theta$$

$$\nabla J(\theta) = 2(X^T Y) + 2X^T X\theta$$

$$\nabla J(\theta) = 0 \qquad 2X^T Y = 2X^T X\theta$$

$$(X^T X)^{-1} X^T Y = \theta$$

"Minima"

In practice

We must do.

3 dim

$$\theta^T A\theta + B^T\theta + C$$

double derivative

Hessian

$(A > 0)$
$(B \geq 0)$
$(C > 0)$

$(A < 0)$
$(B < 0)$

$ax^2 + bx + c$

X: feature matrix

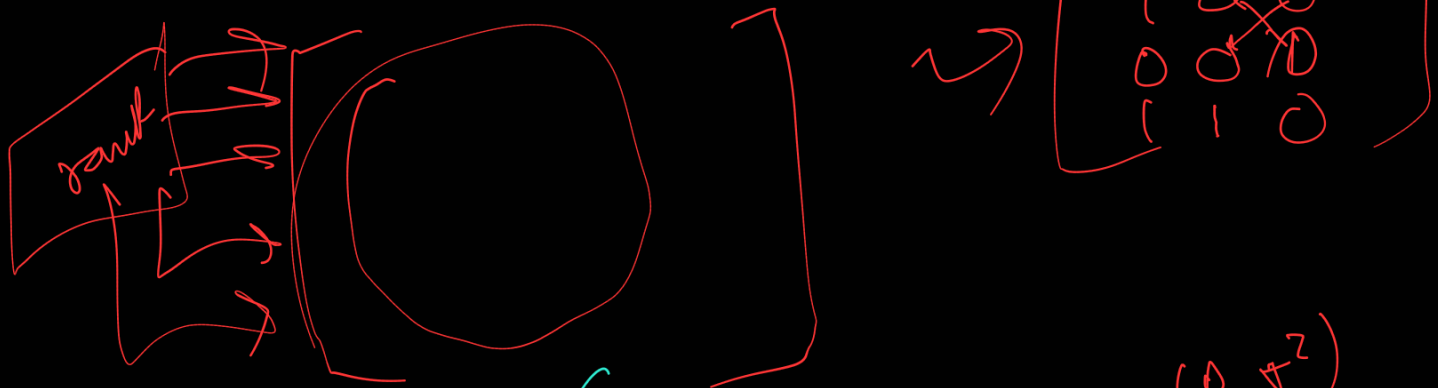
$(x^{(i)})^T x^{(i)} \geq 0$

X: $m \times n$

$x^{(i)} = n \times 1$

$(m \times n)$ $(n \times m)$

$(n \times n)$

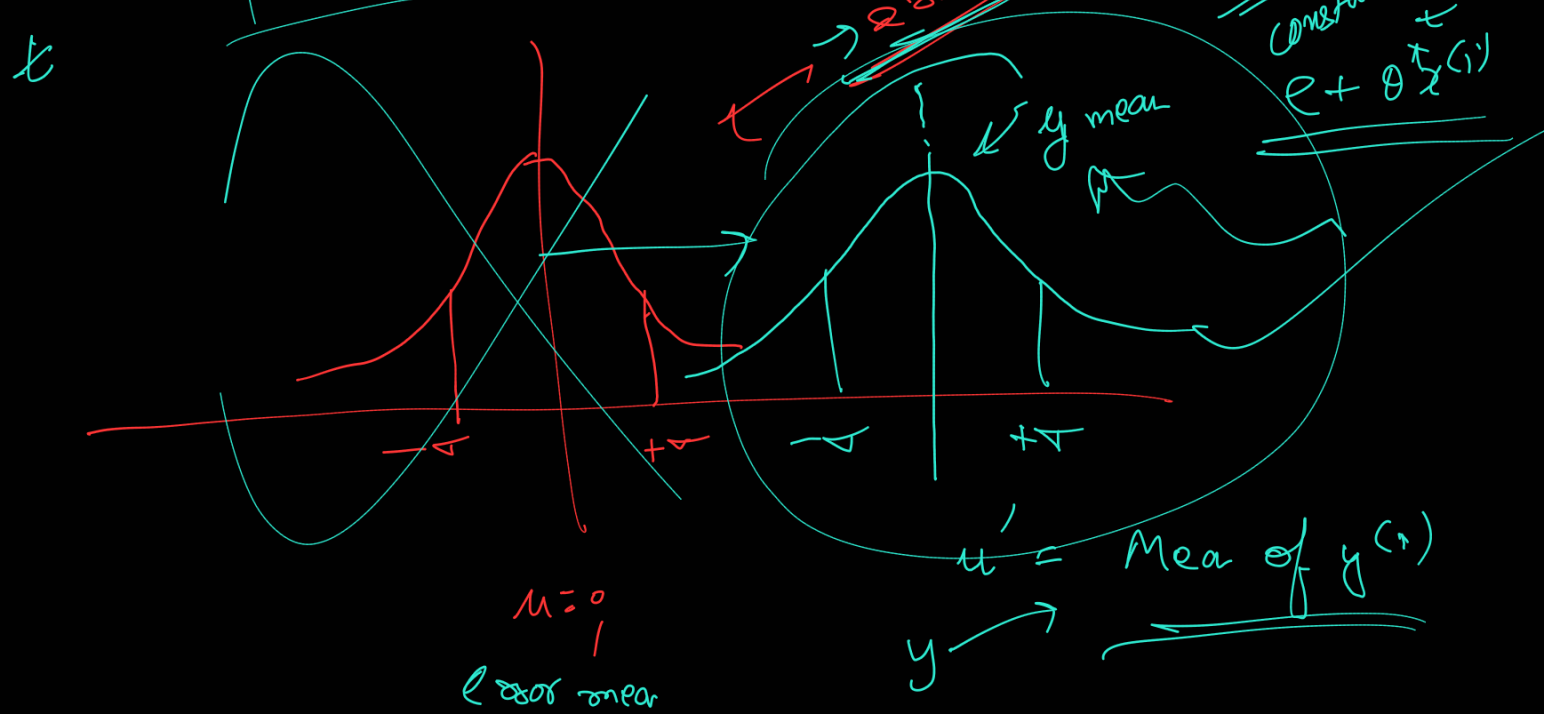


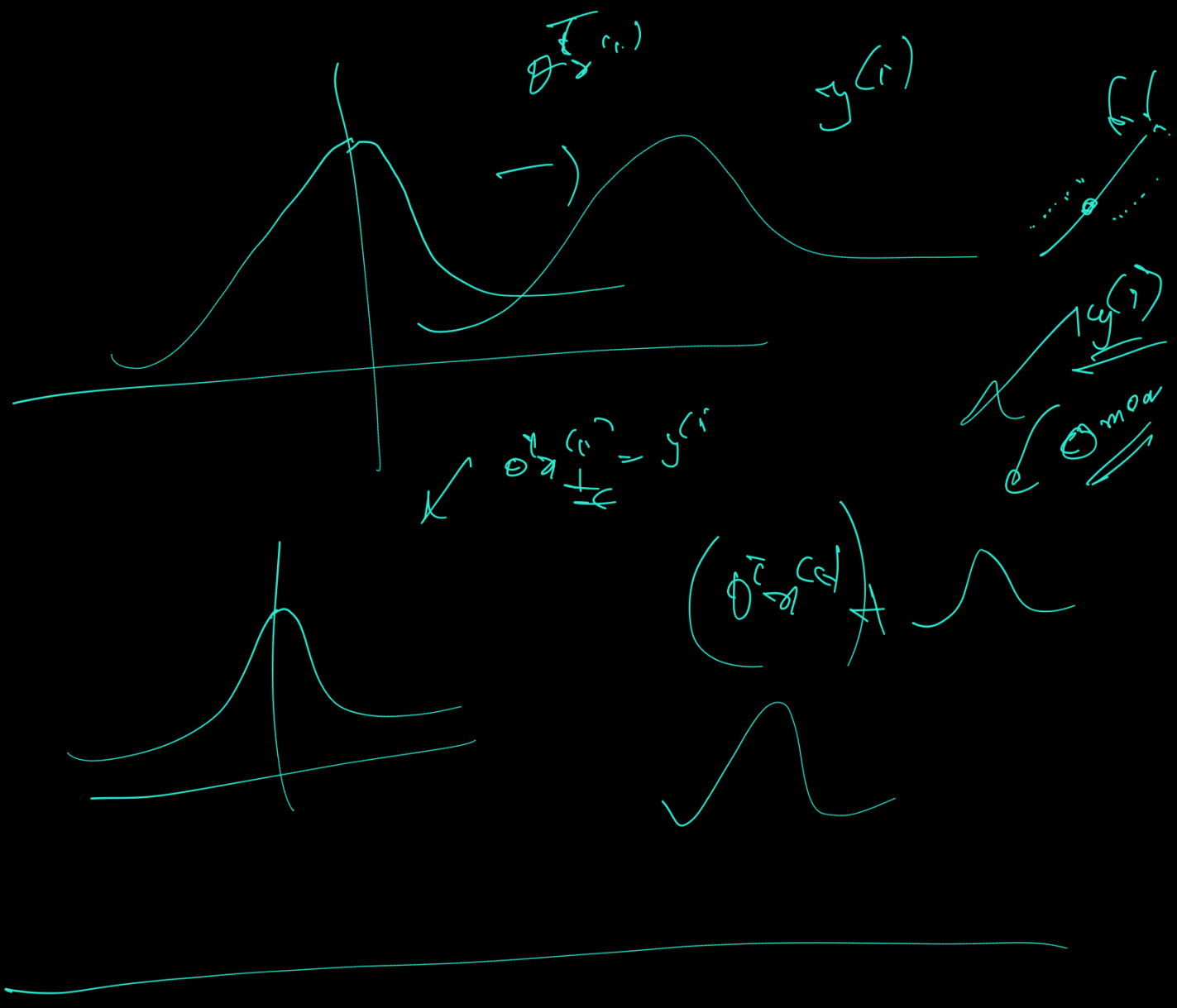
$y^{(i)} = \theta^T x^{(i)} + \epsilon$

$\epsilon \sim N(0, \sigma^2)$

$\theta^T x^{(i)}$ mean of constant value

$\epsilon + \theta^T x^{(i)}$




$\theta^T x^{(i)}$
$y^{(i)}$
$y^{(i)}$
$\theta^T x^{(i)} = y^{(i)}$
$(\theta^T x^{(i)})$