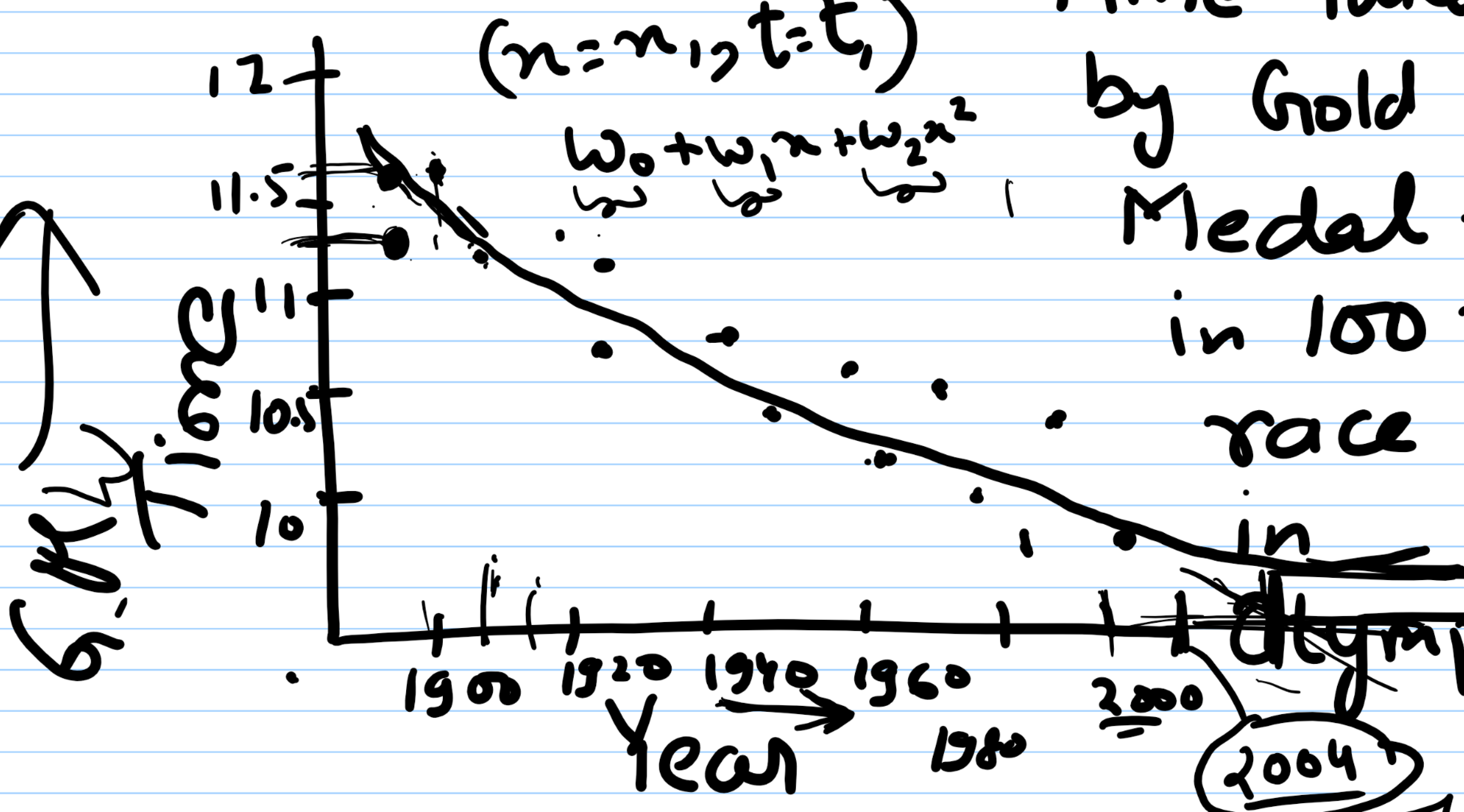
Regression: Fitting an Model on a data Points.

Time taken by Gold Medal in 100 m race in Olympics



eq of family of line

$$y = mx + c$$

eq^n of a line

$$y = m_1 x + c_1$$

t	x
t_1	x_1
t_2	x_2
$\vdots$	$\vdots$
t_n	x_n

Say for $x = x_1$

Years $\Rightarrow x \rightarrow$ independent Variable

time $\Rightarrow y \rightarrow$ dependent variable.

$$y = m_1 x + c_1$$

$$m_1 \,\&\, c_1 \in \mathbb{R}$$

Correspoint to $x = x_1$

$$y_1 = m_1 x_1 + c_1$$

$\Rightarrow$ from the model.

If True = The Value from the Model.

Actual t_1

$t_1 = y_1$ = Estimated Value.

Usually if it is not so.

If the model outcome is not equall to True value than we can say there an error.

t	x
t_1	x_1
t_2	x_2
t_3	x_3
$\vdots$	$\vdots$
t_n	x_n

$n \rightarrow$ Size of the dataset.

Actual true

from the model

$$\text{error} = (t_1 - y_1)$$

$$\text{Total} = (t_1 - y_1) + (t_2 - y_2) + \cdots + (t_n - y_n)$$

Totalerror = $\sum_{i=1}^{n}(t_i - y_i)$

$$t_i - y_i > 0$$

$$t_i - y_i < 0$$

$$t_1 - y_1 = 1000$$

$$t_2 - y_2 = -1000$$

$$t_3 - y_3 = +500$$

$$t_4 - y_4 = -500$$

$$y_1 = m x_1 + C$$

$$\text{error} = \| t_i - y_i \|$$

$$\text{Total error} = \sum_{i=1}^{n} \| t_i - y_i |$$

$$\text{Total error} = \sum_{i=1}^{n} | t_i - (mx_i + c) |$$

$$\underline{\text{Total error}} = \underline{\sum_{i=1}^{N}} \underbrace{\left(t_i - (\underbrace{mx_i + c}) \right)^2}$$

as the deviation increas from the true value.

the function value will also increase.

$$\text{Total Error} = \sum_{i=1}^{n} \left(t_i - (mx_i + c)\right)^2$$

$$\underline{f(\cdot)} = \sum_{i=1}^{n} \left(t_i^2 + (mx_i + c)^2 - 2t_i(mx_i + c)\right)$$

$$\underline{f(m, c)} = \sum_{i=1}^{n} \left(t_i^2 + m^2 x_i^2 + c^2 + 2mx_ic - 2mt_ix_i - 2t_ic\right)$$

$$2c \sum_{i=1}^{N} 1 = N$$

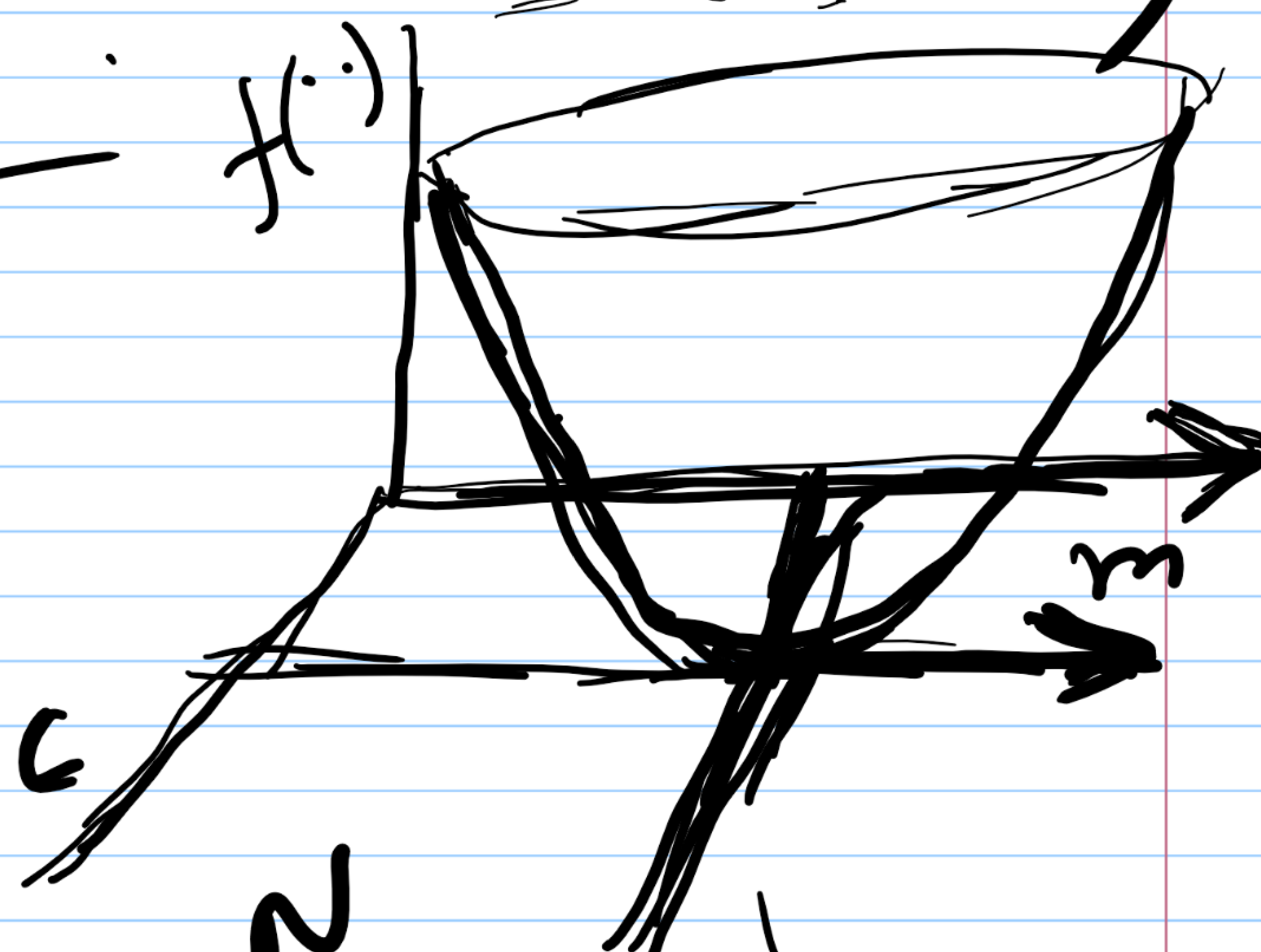
$$\frac{\partial f}{\partial m} = \frac{\partial}{\partial m} \sum_{i=1}^{n} \left(t_i^2 + m^2 x_i^2 + c^2 + 2mx_ic - 2mt_ix_i - 2t_ic\right)$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial m} \left(t_i^2 + m^2 x_i^2 + c^2 + 2mx_ic - 2mt_ix_i\right.$$

$$-2t_i c)$$

$$\frac{\partial f}{\partial m} = \sum_{i=1}^{n}\left(0 + 2mx_i^2 + 0 + 2x_i c - 2t_i x_i - 0\right)$$

$$\frac{\partial f}{\partial m} = \left(\sum_{i=1}^{n}\left(2mx_i^2 + 2x_i c - 2t_i x_i\right)\right)$$



$$\frac{\partial f}{\partial m} = 0$$

$$\sum_{i=1}^{N} m x_i^2 + \sum_{i=1}^{N} x_i c - \sum_{i=1}^{N} t_i x_i = 0 \quad (1)$$

$$\boxed{m\sum_{i=1}^{N} x_i^2 + c\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} t_i x_i = 0} \quad (1)$$

$$\frac{\partial f}{\partial c} = 0 + 0 + 2cN + 2m\sum_{i=1}^{N} x_i + 0 - 2\sum_{i=1}^{n} t_i$$

$$\frac{\partial f}{\partial c} = 0$$

$$\underline{2cN} + \underline{2m}\sum_{i=1}^{N} x_i - 2\sum_{i=1}^{N} t_i = 0$$

$$2N\left(c + m\underline{\frac{1}{N}\sum_{i=1}^{N} x_i} - \frac{1}{N}\sum_{i=1}^{N} t_i\right) = 0$$

$$c + m\bar{x} - \bar{t} = 0$$

$$c = \bar{t} - m\bar{x} \quad \text{--- (2)}$$

$$m\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} t_i x_i = 0 \quad \text{--- (1)}$$

divide eq (1) by N.

$$m\overline{x^2} + c\bar{x} - \overline{tx} = 0 \quad \text{--- (1)}$$

$$m\overline{x^2} + (\underline{\bar{t} - m\bar{x}})\bar{x} - \overline{tx} = 0$$

$$m\left(\overline{x^2} - (\bar{x})^2\right) = \overline{tx} - \bar{t}\,\bar{x}$$

$$\Rightarrow \boxed{\hat{m} = \frac{\overline{tx} - \bar{t}\,\bar{x}}{\overline{x^2} - (\bar{x})^2}}$$

$$\Rightarrow \boxed{\hat{c} = \bar{t} - \hat{m}\,\bar{x}}$$

$$\boxed{y = \hat{m}x + \hat{c}}$$

corresponding to which the error is minimum when the model is linear.

$\Rightarrow \overline{tx} = 20268.1$

$\Rightarrow \overline{x^2} = 3.8130 \times 10^6$

$\Rightarrow \bar{x} = 1952.37$

$$\Rightarrow \bar{t} = 10.39$$

$$\Rightarrow \hat{m} = \frac{\overline{tx} - \bar{t} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2}$$

$$= \frac{20268.1 - 10.39 * 1952.37}{3.8130 \times 10^6 - (1952.37)^2}$$

$$= \frac{-1.0243}{125.3831}$$

$$\boxed{\hat{m} = -0.0136}$$

$$\hat{c} = \bar{t} - \hat{m}\bar{x}$$

$$= \underline{10.39} + \underbrace{0.0136}\times(1952.37)$$

$$\hat{c} = 36.95 \Leftarrow$$

$$\boxed{y = -0.0136x + 36.95} \Leftarrow$$

$$y = w_0 + w_1 x + w_2 x^2$$

$$\text{error} = \sum_i (t_i - y_i)^2$$

$$f(w_0, w_1, w_2) = \sum_i \left(t_i - (w_0 + w_1 x + w_2 x^2)\right)^2$$

$$\frac{\partial f}{\partial w_0}, \frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}$$

⇒ Total error corresponding to 2^n order model <

Total error corresponding to 1^{rd} order model.

Go for Matrix/Vector representation of the model

$$\boxed{y_i = \underline{w}^T \underline{x}_i} \Leftarrow$$ corresponding to i^{th} data point

$$\Rightarrow \underline{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}_{2\times 1}, \quad x_i = \begin{bmatrix} 1 \\ x_i \end{bmatrix}_{2\times 1}$$

$$y_n = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

$$= \begin{bmatrix} w_0 & w_1 \end{bmatrix}_{1\times 2} \begin{bmatrix} 1 \\ x_i \end{bmatrix}_{2\times 1}$$

$$x_i = \begin{bmatrix} 1 \\ x_i \\ x_i^2 \\ \vdots \\ x_i^n \end{bmatrix}_{(n+1)\times 1}$$

$$\Rightarrow y_{i\,1\times 1} = \begin{bmatrix} w_0 + w_1 x_i \end{bmatrix}_{1\times 1}$$

$$\Rightarrow \underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}_{(n+1)\times 1} - w^T x_i = w_0 + w_1 x_i + w_2 x_i^2 + \cdots + w_n x_i^n$$

$$\Rightarrow \alpha = \frac{1}{N} \boxed{\sum_{i=1}^{N} \Big(t_i - \underbrace{w^T x_i}_{y_i} \Big)^2}$$

Average error

Vector representation of the same is

$$\Rightarrow \boxed{\alpha = \frac{1}{N} (\underset{\sim}{t} - \underset{\sim}{X} w)^T (t - Xw)}$$

where matrix $\underline{X}$ is combination of all x_i

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_N \end{bmatrix}_{N \times 2}$$

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

$$\begin{bmatrix} 1 & x_i \end{bmatrix}$$

$$t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}_{N\times 1} \qquad X = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_N & x_N^2 & & x_N^n \end{bmatrix}}_{N\times(n+1)}$$

$$Xw = \underline{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_N \end{bmatrix}}_{\underline{N\times 2}} \times \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}_{\underline{2\times 1}}$$

$$Y = \underline{X\,w} = \begin{bmatrix} w_0 + w_1 x_1 \\ w_0 + w_1 x_2 \\ \vdots \\ w_0 + w_1 x_N \end{bmatrix}_{\underline{N\times 1}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

General case when $n = n$.

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_N & x_N^2 & \cdots & x_N^2 \end{bmatrix}_{N\times(n+1)}$$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}_{(n+1)\times 1}$$ ✓

$$Xw = \begin{bmatrix} & \\ & \end{bmatrix}_{\underline{N\times(n+1)}} * \begin{bmatrix} & \\ & \end{bmatrix}_{\underline{(n+1)\times 1}}$$

$$= \begin{bmatrix} & \\ & \end{bmatrix}_{\underline{N\times 1}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Avg Error

$$\alpha = \frac{1}{N} \boxed{(t - Xw)^T (t - Xw)}$$

$(Xw)^T$
$= w^T X^T$

$$= \frac{1}{N}\left(t^T - w^T X^T\right)\left(t - Xw\right)$$

$$\mathcal{L} = \frac{1}{N}\left(t^T t - w^T X^T t - t^T X w + w^T X^T X w\right)$$

$$\mathcal{L} = \frac{1}{N}\left(t^T t - 2 w^T X^T t + w^T X^T X w\right)$$

$w^T X$

$$\because \left(w^T X^T t\right) = \left(t^T X w\right)^T = (Xw)^T t = w^T X^T t$$

Order of $t^T X w$ is $(t^T)_{1 \times N} * (X)_{N \times (n+1)} * (w)_{(n+1) \times 1}$

$$= [\quad]_{1 \times 1}$$

$[2]^T = [2]$

$[a]^T = (a)$

$$\Rightarrow \left[t^T X w\right]_{1 \times 1}$$

$$\left[\left[\quad \right]_{1\times 1} \right]^T = \left[\quad \right]_{1\times 1}$$

If the matrix have single element then transpose of the matrix same as the original matrix.

Rules for derivative identities with respect to a vector.

$f(w)$	$\frac{\partial f}{\partial w}$
$w^T x$	x
$x^T w$	x
$w^T w$	$2w$
$w^T C w$	$2Cw$

$$\Rightarrow \frac{\partial L}{\partial \underline{w}} = 0 + \frac{2}{N} X^T X \underline{w} - \frac{2}{N} X^T t$$

$$= 0$$

$$\Rightarrow X^T X \underline{w} = X^T t$$

$$\Rightarrow \boxed{w = (X^T X)^{-1} X^T t}$$

$$(X^T X)^{-1} (X^T X) w = (X^T X)^{-1} X^T t$$

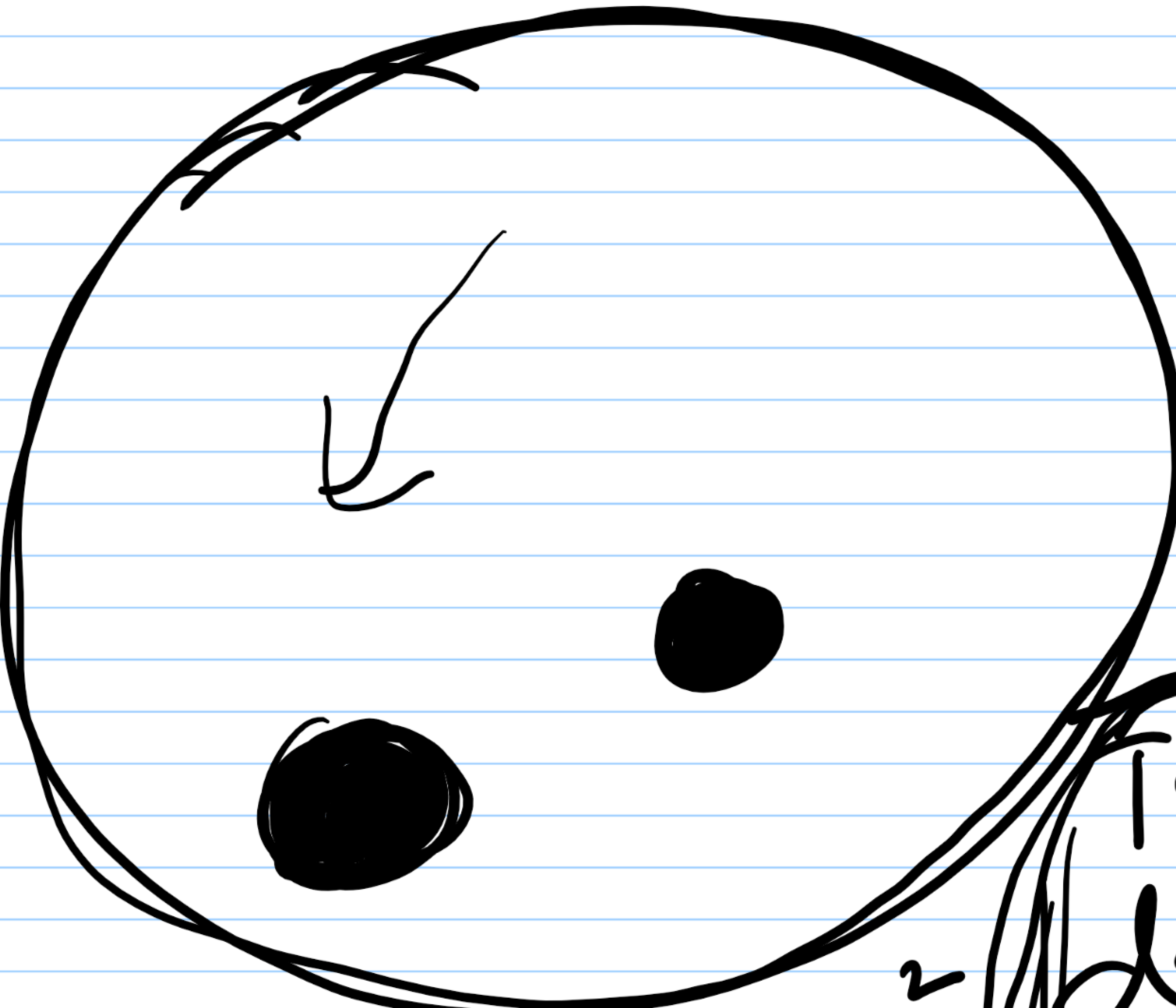
$A^{-1}A = I$

$IA = A$

$$I w = (X^T X)^{-1} X^T t$$

$$\boxed{w = (X^T X)^{-1} X^T t}$$

$$\underline{w} = \begin{bmatrix} \vec{w_0} \\ \vec{w_1} \\ \vdots \\ \vec{w_n} \end{bmatrix}$$



data

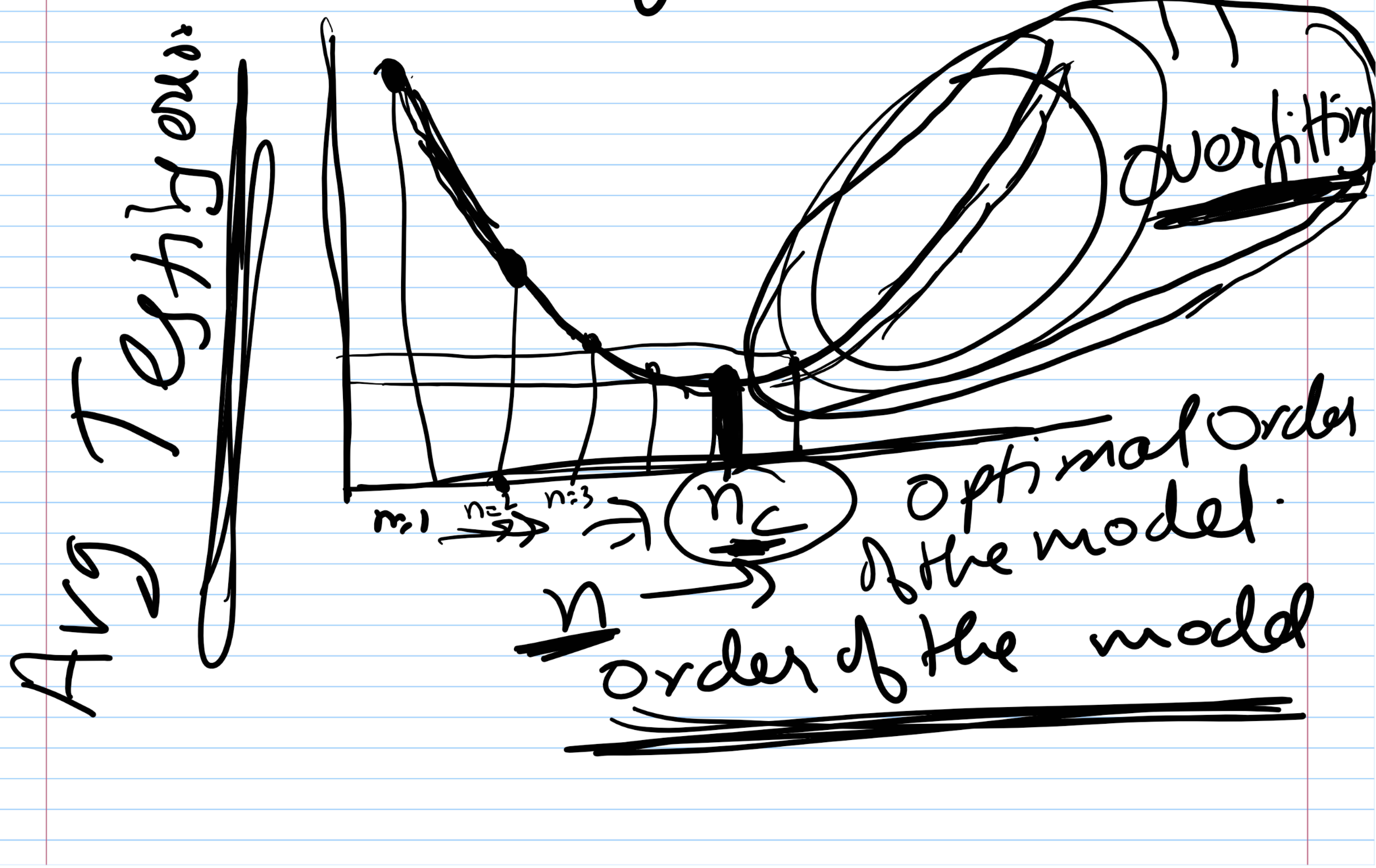
90% Training data set

10% Testing data set

$\Rightarrow Y = Xw$

$\Rightarrow Error_{testing} = \frac{1}{N}\sum (t_i - y_i)^2$

Total testing error



$$f(x) = w_0 + w_1 x + w_2 x^2$$

Traing dataset

x	t
x_1	t_1
$\vdots$	$\vdots$
x_n	t_n

$$\text{Avg error} = \frac{1}{N}\sum_i \left(t_i - f(x_i)\right)$$

n is large