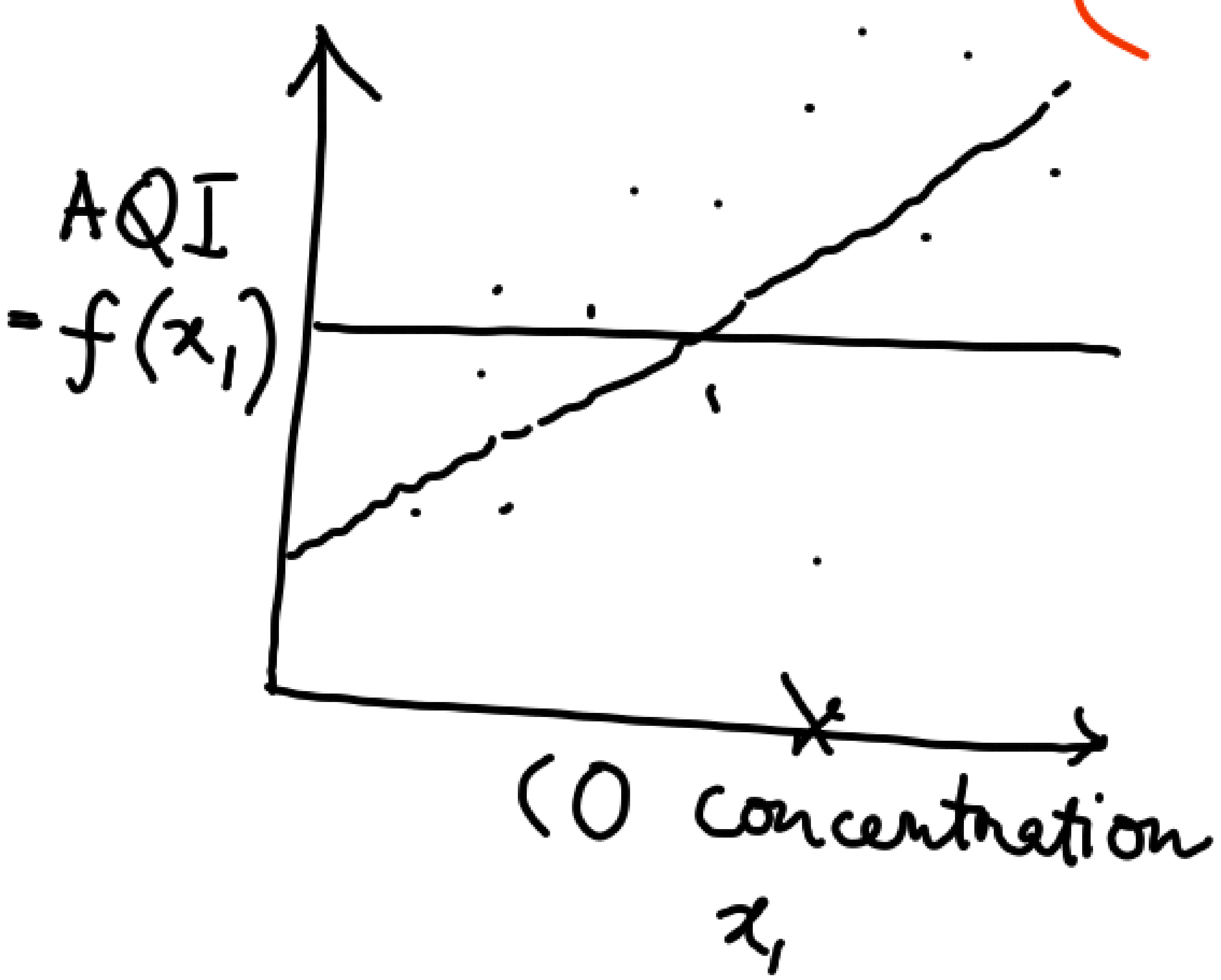
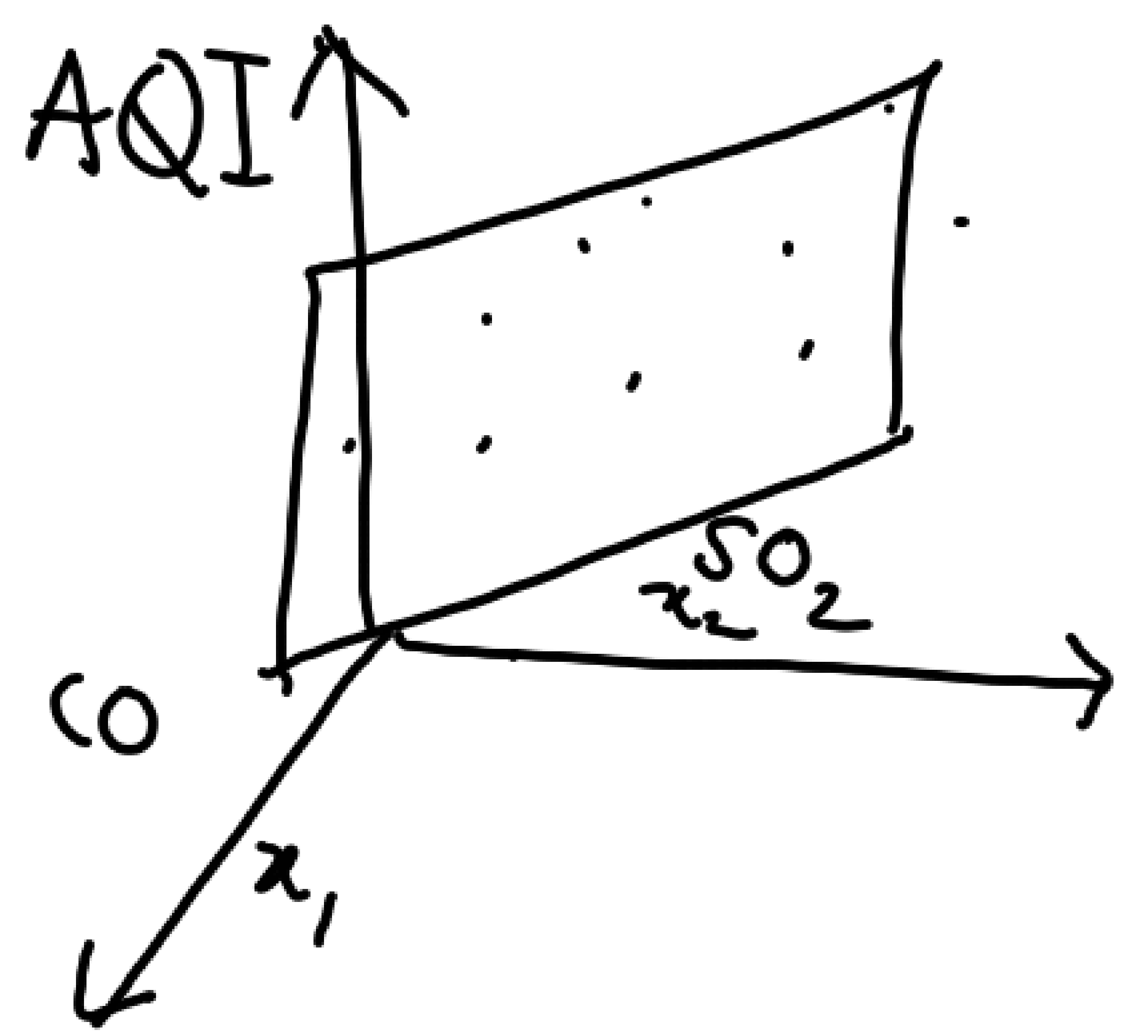
Lec 3 : Regression

Ex! AQI : Air Quality Index : PM2.5 PM10 CO, SO_2, . . . O_3

$$AQI = \max\{f_1(x_1), f_2(x_2), \ldots, f_n(x_n)\}$$



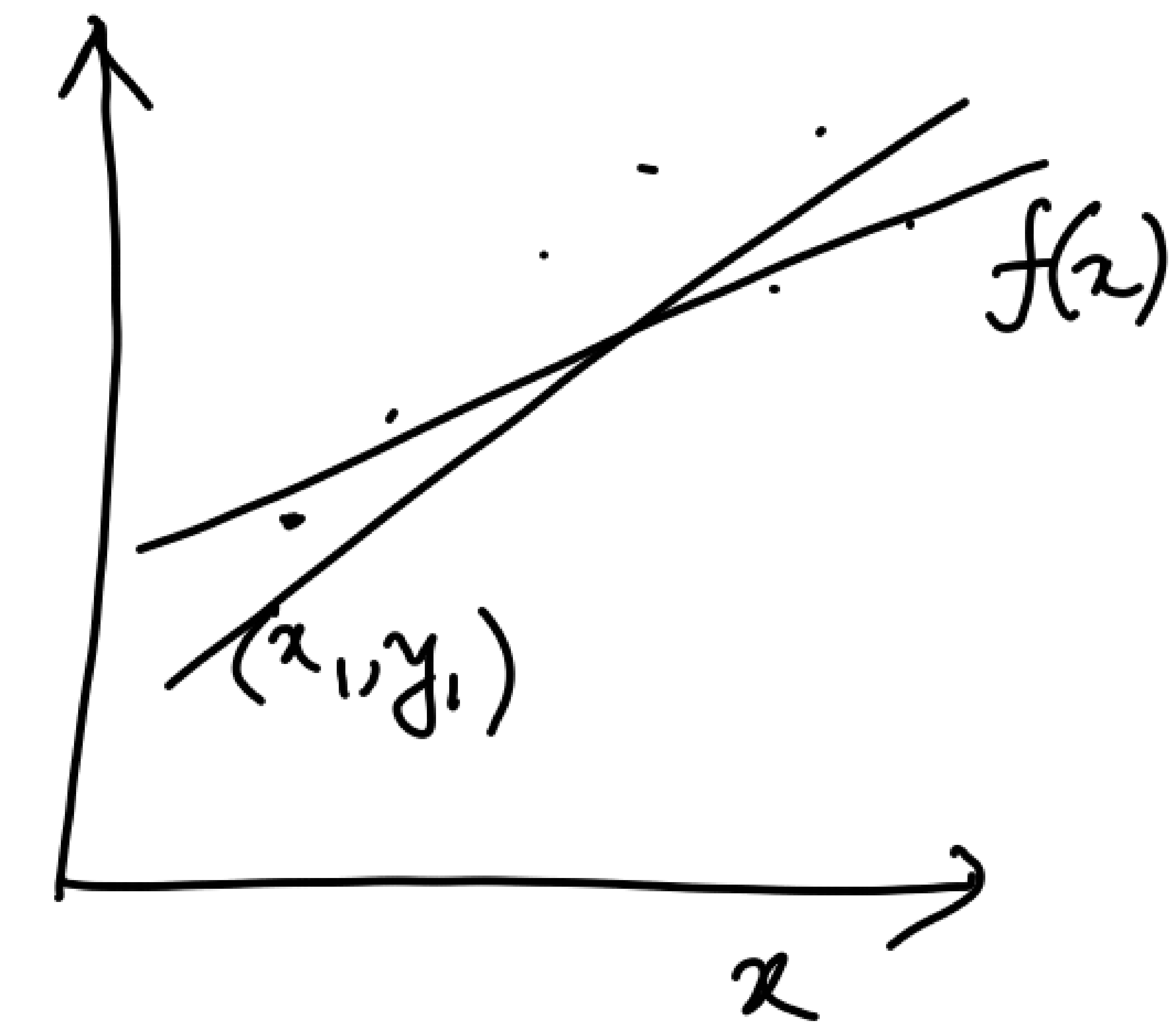


→ function of various pollutant concentrations

→ One plausible way is to fit the perfect AQI data with limited observation and estimate AQI

Linear Regression

- Simplicity but powerful tool
- Interpretable
- Works on transformations of raw data



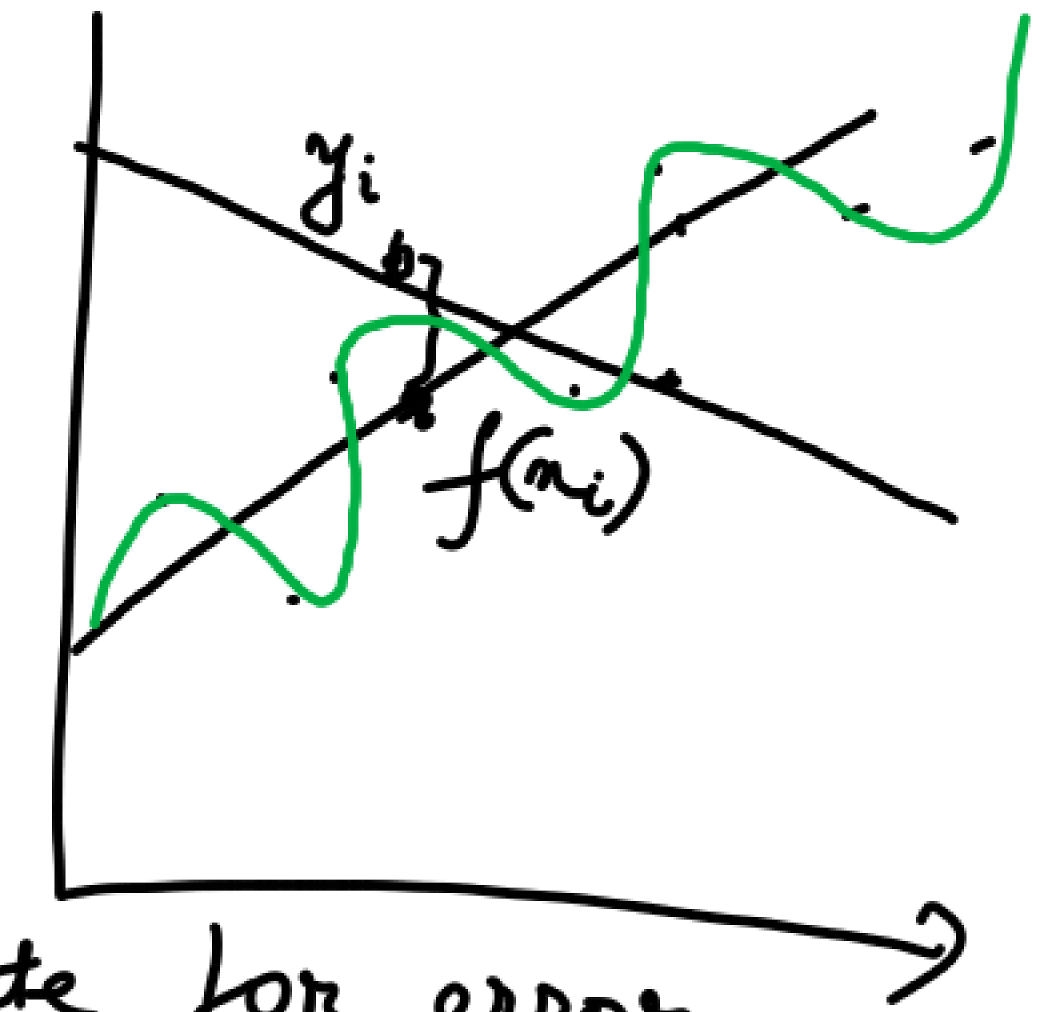
Q: How to best fit the given data?

Measure the goodness of the fit using an error function:

Error
loss
cost
energy

$$E(f, D) \qquad \underset{\text{dataset}}{D} = \{(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)\}$$

Possible E functions:



① $$\sum_{i=1}^{n}\left(f(x_i) - y_i\right)$$

signed, not a good candidate for error

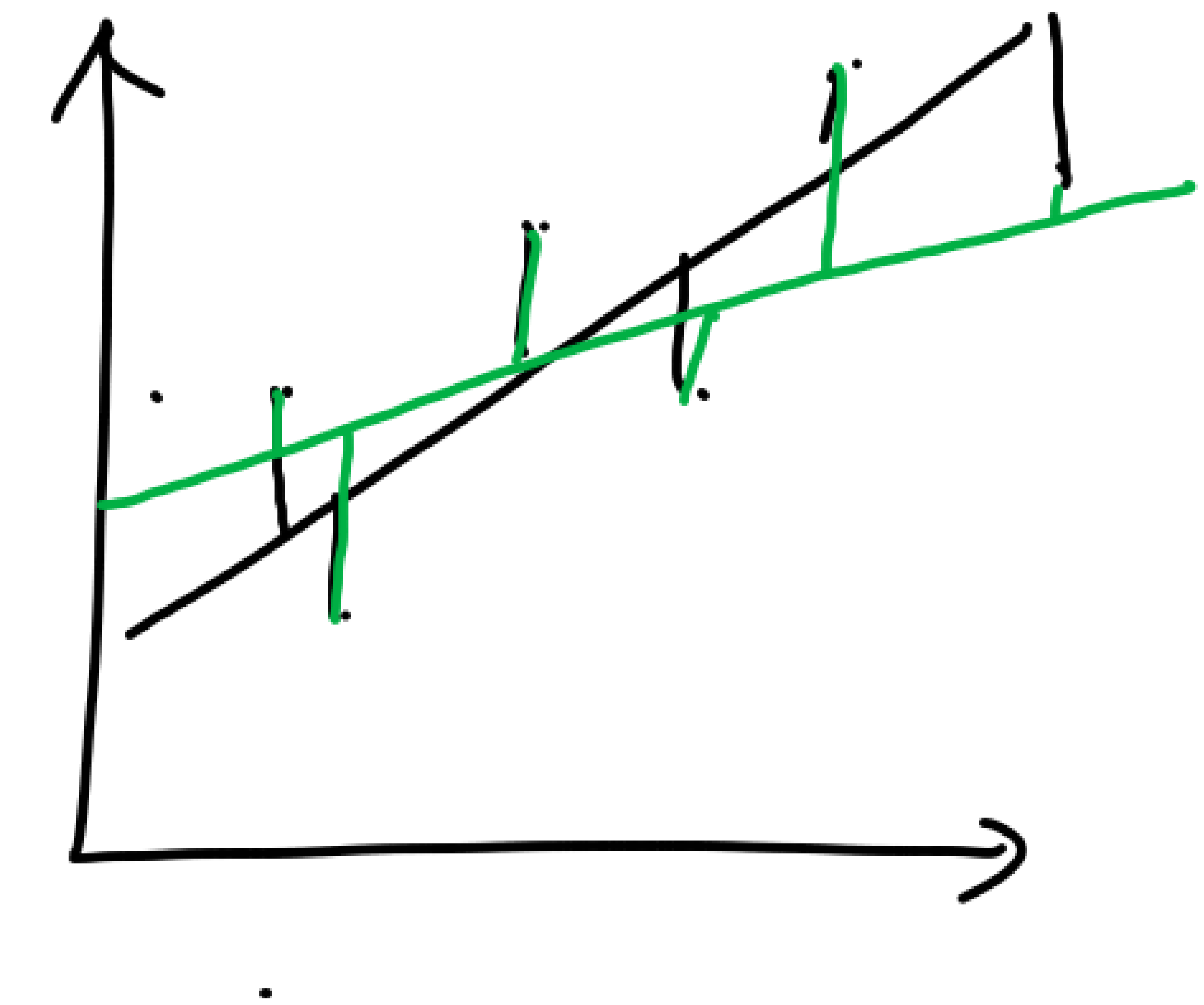
② $$\sum_{i=1}^{n}\left|f(x_i) - y_i\right| \rightarrow \geq 0$$ a good candidate

③ $$\sum_{i=1}^{n}\left(f(x_i) - y_i\right)^2$$ (squared error)

④ $$\sum_{i=1}^{n}\left(f(x_i) - y_i\right)^3 \rightarrow \times$$

Squared error function

$$\sum_{\{i:\ x_i, y_i \in D\}} \left(f(x_i) - y_i \right)^2$$



- Continuous, differentiable
- Visualizable in Euclidean space
- Mathematical analysis becomes easier.

input → output

$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ $\quad (x_j, y_j)$: j^{th} training example.

n = # of data samples / training instances

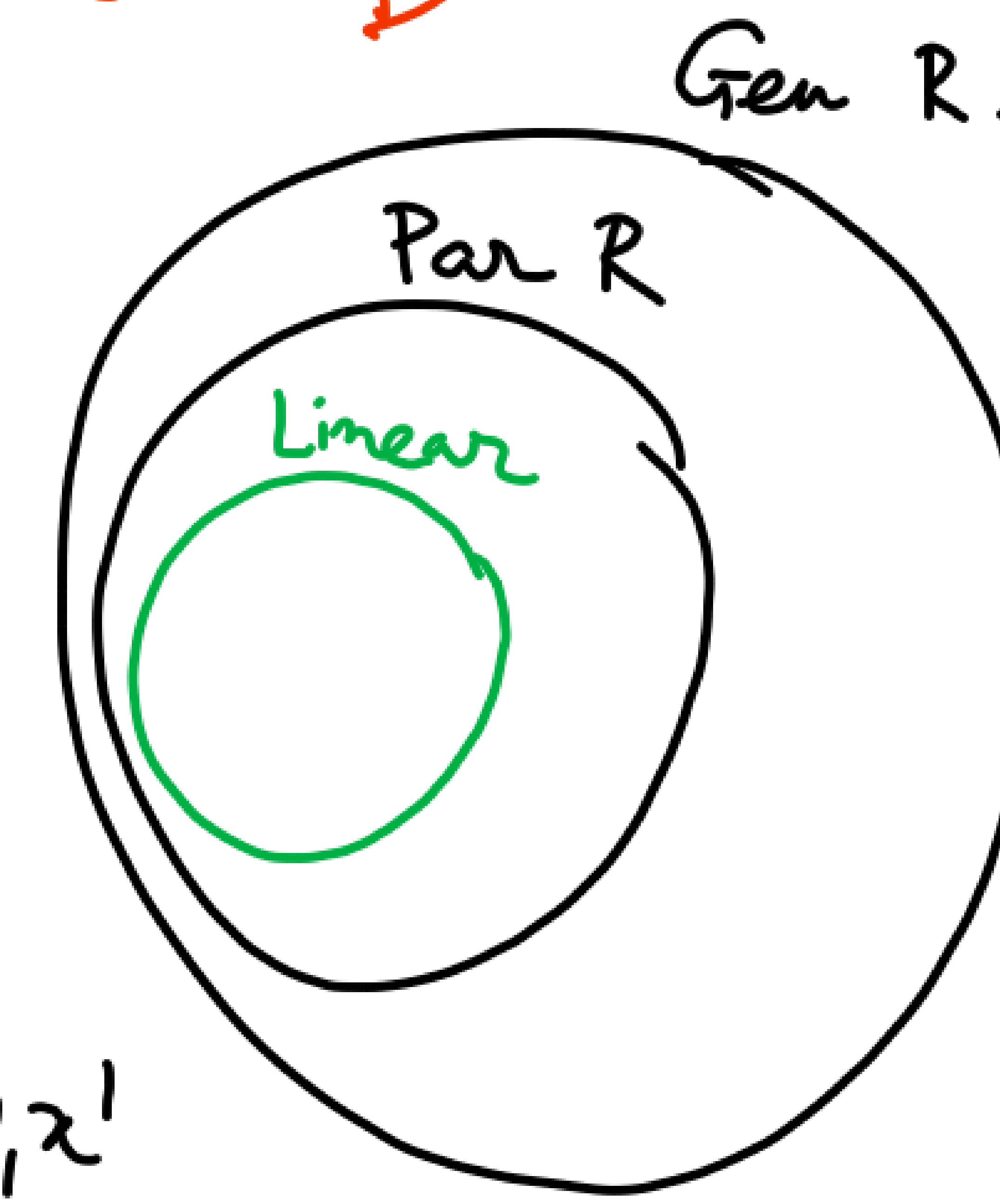
$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} \begin{matrix} \rightarrow PM2.5 \\ \rightarrow CO \\ \rightarrow SO_2 \end{matrix} \qquad x_i \in \mathbb{R}^d \qquad y_i \in R$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & & \\ x_{n1} & \cdots & & x_{nd} \end{bmatrix} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}_{n \times d} ; \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

General Regression

find a fn. f^* s.t. $f^*(x)$ is the best predictor of y w.r.t. D

$$f^* \in \underset{f \in \mathcal{F}}{\operatorname{argmin}} E(f, D)$$



Parametrized Regression

$$f \equiv f(x, w)$$

$$\underset{w}{\operatorname{argmin}} E(f(x,w), D)$$

Ex: 1. $f(x, (\alpha, \lambda)) = \alpha e^{-\lambda^T x}$

2. $f(x, w) = w_0 + w_1 x^1 + \dots + w_k x^k$

Linear Regression

$$f(x, w) = \underline{W^T x + w_0} = \bar{W}^T x$$

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

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}_{d \times 1} = W \qquad x_i \in \mathbb{R}^d$$

$$f(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m$$

Least square optimization for lin regression

$$w^* \in \underset{w}{\text{argmin}} \sum_{i=1}^{n} \left(\sum_{j=1}^{d} w_j x_{ij} - y_i \right)^2$$

$$\boxed{\hat{y}_i = \sum_{j=1}^{d} w_j^* x_{ij}}$$

$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$

$\underline{d=1}$: $E(\underline{w}, D) = \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2$

Find w_0, w_1 s.t. $\frac{\partial E}{\partial w_0} = 0 \quad \frac{\partial E}{\partial w_1} = 0$

$$\frac{\partial E}{\partial w_0} = 0 \Rightarrow -2 \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i) = 0$$

$$\Rightarrow w_0 = \frac{\sum y_i - w_1 \sum x_i}{n} = \bar{y} - w_1 \bar{x}$$

$$\frac{\partial E}{\partial w_1} = 0 \Rightarrow w_1 = \frac{\sum x_i y_i - w_0 \sum x_i}{\sum x_i^2}$$

$$\alpha = \frac{\sum x_i y_i}{\sum x_i^2}, \quad \beta = \frac{\sum x_i^2}{n}$$

$$w_1 = \frac{\alpha\beta - \bar{x}\bar{y}}{\beta - \bar{x}^2} \overset{HW}{=} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

<u>d-D data LR</u>

$$x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{id} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$$

$$w^* \in \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \underbrace{\left(y_i - w^T x_i\right)^2}_{=: z_i} = \sum_{i=1}^{n} z_i^2$$

$$= \|z\|^2$$

$$= \|y - Xw\|^2$$

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} y_1 - x_1^T w \\ y_2 - x_2^T w \\ \vdots \\ y_n - x_n^T w \end{bmatrix} = y - Xw$$

$$\underset{w}{\mathrm{argmin}}\ (Xw - y)^T (Xw - y) \quad = \quad \|y - Xw\|^2$$

$$E(w, D) = w^T \underline{X^T X} w - 2\underline{y}^T X w + y^T y$$

$$\nabla_w E = 0 \Rightarrow \quad 2X^T X w - 2X^T y = 0$$

$$\frac{\partial E}{\partial w_1} = \qquad \Rightarrow \boxed{w^* = (X^T X)^{-1} X^T y} \qquad \boxed{w^{*T} x = \hat{y}}$$

Geometry representation : Next time