Lecture 11 – Curve fitting

Curve fitting → discrete data → continuous function

↓

find values where not available

Approximation of functions

→ regression → to go back to the root

→ Interpolation → "inter" "polire" → Latin

b/w, among polish smooth

$x_i, y_i \rightarrow y_i = f(x_i)$

?

$y_i \longrightarrow y^*$

y_{i-1} $\overset{?}{y_i}$ y_{i+1}

given x_i, y_i find $f(x_i)$ such that the <u>discrepancy</u>

between y_i & $f(x_i)$

$$e_i = y_i - f(x_i) \quad \text{is minimal}$$

Strategies

$\bar{y}$

Problem

$\bar{y}$ y_i - midpoint

$$\sum e_i = \sum (y_i - f(x_i))$$

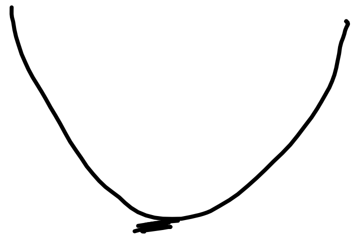
minimize the sum of the <u>errors</u>

$$\sum |e_i| = \sum |y_i - f(x_i)|$$

→ if you have prior information → constrained solution

n - number of solutions

non-uniqueness



$$\sum e_i^2 = \sum (y_i - f(x_i))^2$$

Cost-function $\leftarrow \underbrace{\sum e_i^2}_{S} = \sum \underline{(y_i - f(x_i))^2} \rightarrow$ Estimation

residuals

$\rightarrow a, b, c, d, e$

Regression

$$\frac{\partial S}{\partial a} = 2(y_i - f(x_i))\left[\cancel{\frac{\partial y_i}{\partial a}} - \frac{\partial f}{\partial a}\right]$$

$$\frac{\partial S}{\partial b}$$

$$\frac{\partial S}{\partial c}$$

$\boxed{\text{Linear}}$ system of equations

$\downarrow$

Least-squares

Non-linear

$$y_i = f(x_i) = f(\tilde{x}_i) + \frac{\partial f}{\partial \tilde{x}_i}(\tilde{x}_i - x) \boxed{+ \ldots}$$

$$\Delta y \leftarrow y - f(\tilde{x}_i) = \frac{\partial f}{\partial \tilde{x}_i}\underbrace{(\tilde{x}_i - x)}_{\Delta x}$$

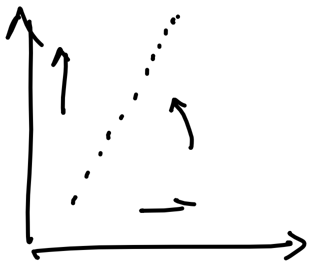
$$\Delta y = \underbrace{\frac{\partial f}{\partial \tilde{x}_i}}_{\text{Jacobian}} \Delta x$$

Linear regression using least squares

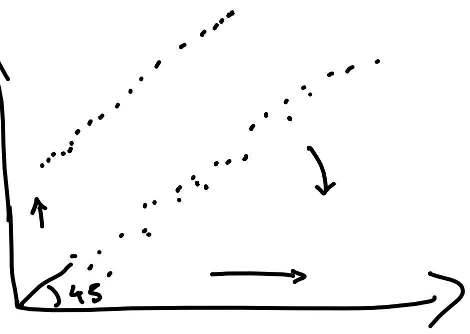
$$y_i = \underbrace{a_0 + a_1 x_i}_{a_0 \to 0,\ a_1 \to 1}$$

$$y_i = x_i$$

new old



new



old

Δx

known

$$S = \sum (e)^2 = \sum (y_i - a_0 - a_1 x_i)^2 \qquad i \in [1, n]$$

not known

$$\frac{\partial S}{\partial a_0} = \sum -2 (y_i - a_0 - a_1 x_i) \to (1)$$

$$\frac{\partial S}{\partial a_1} = \sum -2x_i (y_i - a_0 - a_1 x_i) - (2)$$

(1)

$$\cancel{-2} \sum (y_i - a_0 - a_1 x_i) = 0 - (1a)$$

$$\cancel{-2} \sum x_i (y_i - a_0 - a_1 x_i) = 0 - (2a)$$

(1a) $\Rightarrow$ $\sum y_i = \boxed{\sum a_0} + a_1 \sum x_i = n a_0 + a_1 \sum x_i$

(2a) $\sum x_i y_i = \sum a_0 x_i + a_1 \sum x_i^2$

normal equations

$$\begin{cases} \sum y_i = n a_0 + a_1 \sum x_i \\ \sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \end{cases}$$

$\rightarrow \quad a_0 = \dfrac{\sum y_i - a_1 \sum x_i}{n} = \bar{y} - a_1 \bar{x}$ ($\bar{y}$ → mean, $\bar{x}$ → mean)

→ number of normal equations will always be equal to the number of unknowns

$$\sum x_i y_i = \left(\frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n}\right)\sum x_i + a_1 \sum x_i^2$$

$$n \sum x_i y_i = \sum x_i \sum y_i + n a_1 \sum x_i^2 - a_1 \sum x_i \sum x_i$$

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$$

$\hat{e}_i$ → statistics

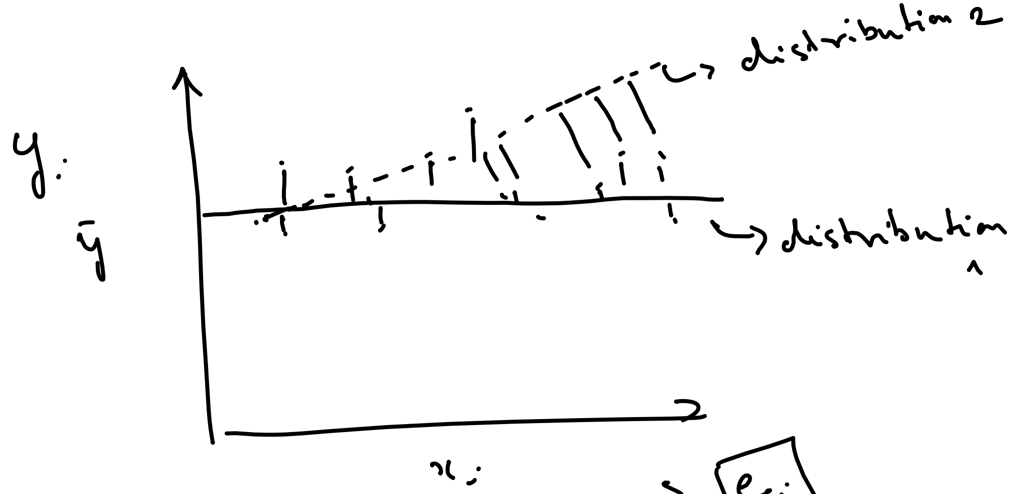
whether $f(x) = a_0 + a_1 x$ is a good model or not?

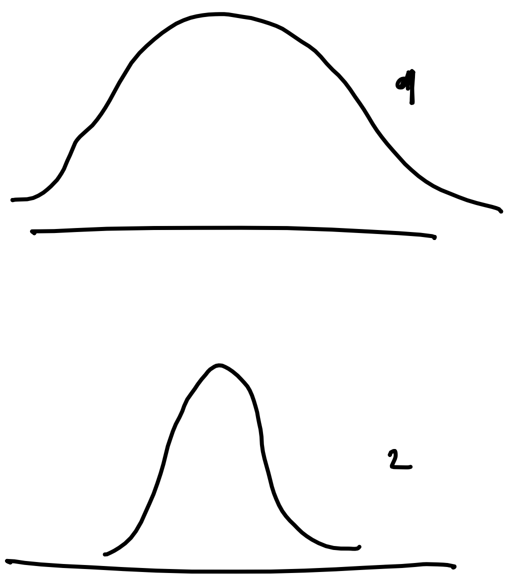
$$S_r = \sum \hat{e}_i^2$$

$$S_t = \sum (y_i - \bar{y})^2$$

$$r^2 = \frac{S_t - S_r}{S_t}$$ → coefficient of determination

S_t





S_r

$\boxed{e_i}$

y_i = Signal + Noise → variability

Signal ↓ variability

Noise ↓ ↳ randomness, stochastics

Noise → precision

S_r

$r^2 \rightarrow \sqrt{r^2} = r \rightarrow$ correlation coefficient

$\downarrow$ $\boxed{0-1}$ $1 \rightarrow -1$

$$r^2 = \frac{S_t - S_r}{S_t} =$$