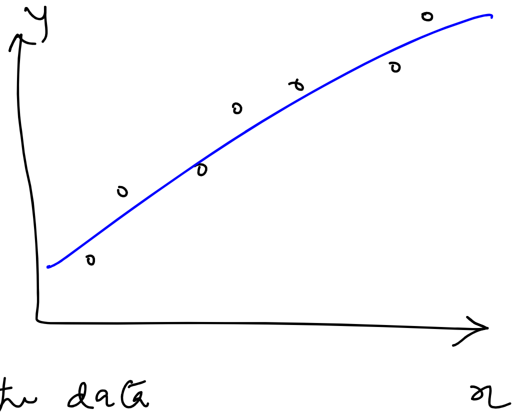
Approximation of Functions

[CURVE FITTING]

Two kinds of problems

(a) Data exhibit a significant degree of scatter

(x_i, y_i) $i = 1, 2 \ldots n$



- Derive a curve that represents the general trend of the data

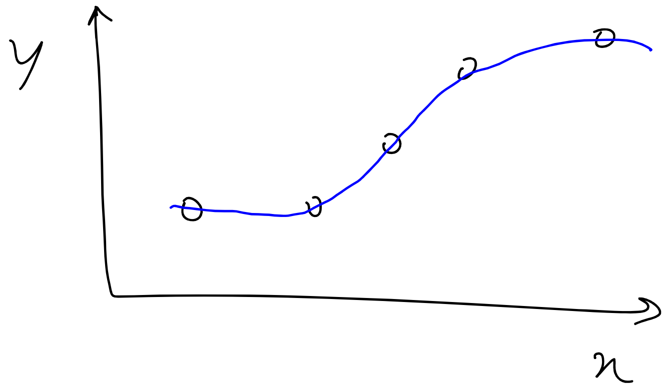
∘ – Approximate fit (Regression)

(b) Data are precise

- Pass the curve or series of curves through each data point

∘ Exact fit

(Interpolation problems)



Regression

- Whether regression or interpolation problem

$\Rightarrow$ WHAT FUNCTION SHALL WE FIT?

It depends on what we want to do!

- Interpolation | Extrapolation
- Integrate
- Differentiation

Question / Answer

Easy to

- Determine
- Evaluate
- Integrate
- Differentiate

Many choices, traditionally

- $\boxed{\text{Polynomials}}$
- Trigonometric functions
- Exponential function
- Other functions depending upon the application

Approximate Fit (Regression)

We will use the Principle of Least Squares

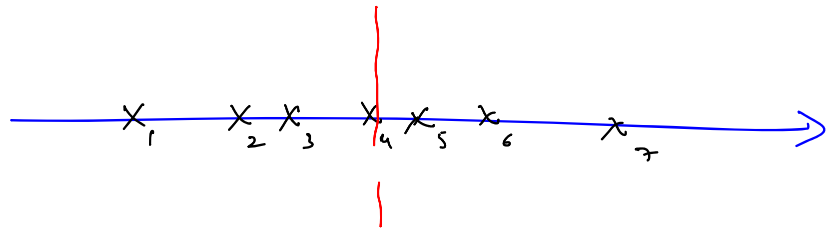
The Principle of Least Squares

Suppose we try to measure a quantity x and make n measurements

x_i, $i = 1, 2, \ldots, n$

We don't know x, but only measurement x_i that have errors e_i

i.e. $x_i = x + e_i$



$x = ?$ $\qquad$ $x = x_{avg}$

The principle of least squares states that the best estimate of x, $\hat{x}$, is that number which minimizes the sum of the squares of the deviations of the data from the estimate.

$$f(\hat{x}) = \sum_{i=1}^{n} (x_i - \hat{x})^2$$

The $\hat{x}$ will be the value which minimizes $f(\hat{x})$

$$f(\hat{x}) = \sum_{i=1}^{n} (x_i - \hat{x})^2$$

$$\frac{df}{d\hat{x}} = -2 \sum_{i=1}^{n} (x_i - \hat{x}) = 0$$

$$\Rightarrow \sum x_i - \sum \hat{x} = 0$$

$$\Rightarrow n\hat{x} = \sum x_i$$

$$\Rightarrow \boxed{\hat{x} = \frac{\sum x_i}{n}}$$

$$\frac{d^2f}{dx^2} = 2n \longrightarrow \text{positive}$$

Here $\hat{x}$ is the minima

Can we have other choices

(a) Minimise absolute error

$$\sum | x_i - \hat{x} | \qquad \hat{x} = x_{median}$$

(b) Minimize the maximum deviation

(Min max)

$$\hat{x} = \frac{x_{min} + x_{max}}{2} = x_{midrange}$$

Chebyshev criteria

Which is the better ?

Regression [Polynomials]

Linear function $y_i = a_0 + a_1 x_i + e_i$

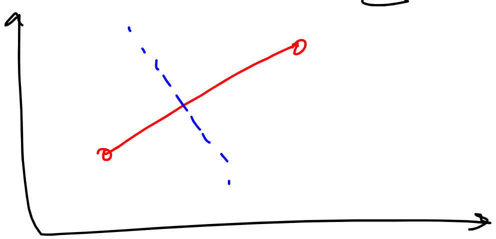
$(x_i, y_i) \quad i = 1, \ldots, n$ $\qquad \hat{y}_i = a_0 + a_1 x_i$

Best choice

(a) Minimize the error (Trivial)

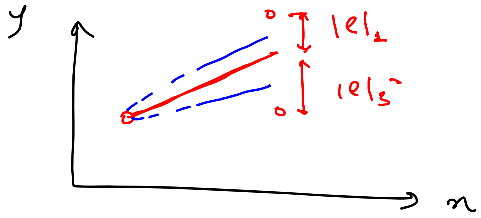
$$\sum e_i = \sum (y_i - \hat{y}_i)$$

$$= \sum (y_i - a_0 - a_1 x_i)$$



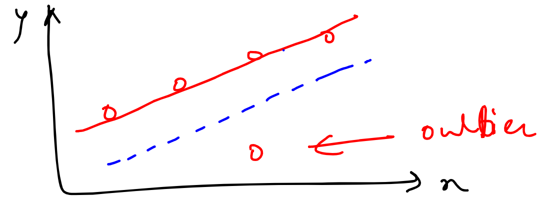
(b) Minimize the absolute error

$$\sum |e_i| = \sum |y_i - a_0 - a_1 x_i|$$



(c) Minimize the maximum error

$$|e_{max}|$$



(d) Minimize sum of squared error

$$\sum e_i^2 = \sum (y_i - a_0 - a_1 x_i)^2$$

Unique solution!

(x_i, y_i) $\hat{y}_i = a_0 + a_1 x_i$

To fit the curve → We need to estimate a_0 & a_1

Least Squares

$$S_r = \sum (y_i - a_0 - a_1 x_i)^2$$

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

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

$$\overbrace{\sum a_0}^{= n a_0} + a_1 \sum x_i = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i$$

$$\Rightarrow \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

normal equations

Remark $y_i = \alpha e^{\beta x_i}$

Linearization of a non-linear equation

$$\ln y_i = \ln \alpha + \beta x_i$$

$$Y_i = a_0 + a_1 X_i$$

$Y_i = \ln y_i$

$\alpha = e^{a_0}$ $\beta = a_1$

Example

$$y = a_0 + a_1 x$$

X	0	2	4	8
y	1	0.7937	0.63	0.3968

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

x	x^2	y	xy
0	0	1	0
2	4	0.7937	1.5874
4	16	0.63	2.520
8	64	0.3968	3.1744
14	84	2.8205	7.2818

$$\begin{bmatrix} 4 & 14 \\ 14 & 84 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 2.8205 \\ 7.2818 \end{bmatrix}$$

$$a_0 = 0.9641$$

$$a_1 = -0.074$$