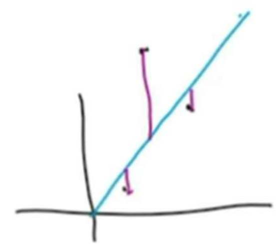
Curve Fitting

Applications of numerical techniques in science & engineering involve curve fitting of experimental data.
If we have experimental data, $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, we want to find a function that can describe the data.
If all numerical values of y_k are known to several significant digits, then polynomial interpolation is used.
But if the data is not available to at least 5 digits of accuracy then we try to find a curve which minimize the error.

The true value of $\overbrace{f(x_k)}^{x_k}$, but the experiment gives y_k.
$f(x_k) = y_k + e_k \Rightarrow e_k = f(x_k) - y_k$

error is measured using different norms.



Maximum error: $E_\infty(f) = \underset{1 \le k \le N}{max} |f(x_k) - y_k|$

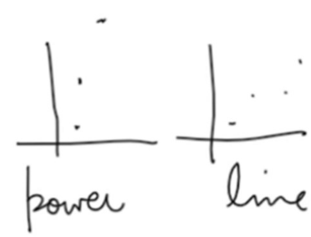
Average error: $E_1(f) = \frac{1}{N} \sum_{k=1}^{N} |f(x_k) - y_k|$

Root mean square error: $E_2(f) = \left[\frac{1}{N} \sum_{k=1}^{N} |f(x_k) - y_k|^2\right]^{\frac{1}{2}}$

The root mean square error $E_2(f)$ is often used when the statistical nature of the error is considered. The curves obtained by minimizing $E_2(f)$ are called Least square curves. The least squares method is the most convenient procedure for **finding best curves.**

We have a **class of functions** that can be used to approximate the data

(Lines, Powers, Exponentials etc.)



Based on the data, we decided which kind of function will be good to approximate the data.
For interpolation polynomial $P_N(x_k) = y_k$ at all nodes x_k.
But for least square curves $f(x_k) \neq y_k$ for all nodes x_{k_1}.

Least Squares Line

We have data (x_k, y_k) for $k = 1, \dots, N$ and we want to find least squares line $f(x) = y = Ax + B$
The root mean squares error is

$$E_2(f) = \left[\frac{1}{N}\sum_{k=1}^{N}|f(x_k) - y_k|^2\right]^{\frac{1}{2}}$$

$$E_2(f) = \left[\frac{1}{N}\sum_{k=1}^{N}|Ax_k + B - y_k|^2\right]^{\frac{1}{2}}$$

$E_2(f)$ will be minimum if $\sum_{k=1}^{N}(Ax_k + B - y_k)^2$ is a minimum.
So, to find the least squares lines we have to minimize
$E_2 = \sum_{k=1}^{N}(Ax_k + B - y_k)^2$

We have to find A & B such that E_2 is a minimum.
The critical points are when $\frac{\partial E_2}{\partial A} = 0$ and $\frac{\partial E_2}{\partial B} = 0$

$$\frac{\partial E_2}{\partial A} = 0 \Rightarrow 2\sum_{k=1}^{N}(Ax_k + B - y_k)(x_k) = 0 \Rightarrow \sum_{k=1}^{N}Ax_k^2 + \sum_{k=1}^{N}Bx_k^2 - \sum_{k=1}^{N}x_k y_k = 0$$

$$\Rightarrow A\sum_{k=1}^{N}x_k^2 + B\sum_{k=1}^{N}x_k = \sum_{k=1}^{N}x_k\, y_k \longleftarrow (1)$$

$$\frac{\partial E_2}{\partial B} = 0 \Rightarrow 2\sum_{k=1}^{N}(Ax_k + B - y_k)(1) = 0 \Rightarrow A\sum_{k=1}^{N}x_k + B\sum_{k=1}^{N}1 - \sum_{k=1}^{N}y_k = 0$$

$$\Rightarrow A\sum_{k=1}^{N}x_k + BN = \sum_{k=1}^{N}y_k \longleftarrow (2)$$

The normal equations for the least squares line are

$$A\sum_{k=1}^{N} x_k^2 + B\sum_{k=1}^{N} x_k = \sum_{k=1}^{N} x_k y_k$$

$$A\sum_{k=1}^{N} x_k + BN = \sum_{k=1}^{N} y_k$$

Solving this linear system we will obtain the values of A & B

Example 1:

Find the least squares line $f(x) = y = Ax + B$ for the data. Also find the error $E_2(f)$.

x_k	-2	-1	0	1	2
y_k	1	2	3	3	4

Solution:

	x_k	y_k	x_k^2	$x_k y_k$
	-2	1	4	-2
	-1	2	1	-2
	0	3	0	0
	1	3	1	3
	2	4	4	8
Sum	0	13	10	7

The normal equation are

$$A\sum_{k=1}^{5} x_k^2 + B\sum_{k=1}^{5} x_k = \sum_{k=1}^{5} x_k y_k \Rightarrow 10A + 0 = 7 \Rightarrow A = \frac{7}{10} = 0.7$$

$$A\sum_{k=1}^{5} x_k + B(\underset{N}{5}) = \sum_{k=1}^{N} y_k \Rightarrow 0 + 5B = 13 \Rightarrow B = \frac{13}{5} = 2.6$$

The least squares line is $f(x) = y = Ax + B = 0.7x + 2.6$

To find the root mean square error $E_2(f) = \left[\frac{1}{N}\sum_{k=1}^{N} |f(x_k) - y_k|^2\right]^{\frac{1}{2}}$

$0.7x_k + 2.6$

In this example, we have

$$E_2(f) = \left[\frac{1}{5}\sum_{k=1}^{5}(0.7x_k + 2.6 - y_k)^2\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{5}\{(0.7(-2) + 2.6 - 1)^2 + (0.7(-1) + 2.6 - 2)^2 + (0.7(0) + 2.6 - 3)^2 + (0.7(1) + 2.6 - 3)^2 + (0.7(2) + 2.6 - 4)^2\}\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{5}\{(0.2)^2 + (-0.1)^2 + (-0.4)^2 + (0.3)^2 + (0)^2\}\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{5}(0.04 + 0.01 + 0.16 + 0.09 + 0)\right]^{\frac{1}{2}}$$

$$= \sqrt{\frac{0.3}{5}} = 0.2449$$

x_k	y_k	$f(x_k) = 0.7x_k + 2.6$	$f(x_k) - y_k$	$(f(x_k) - y_k)^2$
-2	1	$0.7(-2) + 2.6 = 1.2$	0.2	0.04
-1	2	$0.7(-1) + 2.6 = 1.9$	-0.1	0.01
0	3	$0.7(0) + 2.6 = 2.6$	-0.4	0.16
1	3	$0.7(1) + 2.6 = 3.3$	0.3	0.09
2	4	$0.7(2) + 2.6 = 4.0$	0.0	0.00
				0.30

$$E_2(f) = \sqrt{\frac{0.3}{5}} = 0.2449 \; \boldsymbol{ANS}$$

Example 2: Consider the data

x_k	1	2	3	4	5
y_k	1	2	4	6	8

Find the least squares line $y = f(x) = Ax + B$ for the data and also find the root mean square error.

Solution:

	x_k	y_k	x_k^2	$x_k y_k$
	1	1	1	1
	2	2	4	4
	3	4	9	12
	4	6	16	24
	5	8	25	40
Sum	15	21	55	81

The normal equations are

$$A\sum_{k=1}^{5} x_k^2 + B\sum_{k=1}^{5} x_k = \sum_{k=1}^{5} x_k y_k \Rightarrow 55A + 15B = 81 \leftarrow \mathbf{(1)}$$

$$A\sum_{k=1}^{5} x_k + B\underset{\substack{\uparrow \\ \text{N}}}{(5)} = \sum_{k=1}^{5} y_k \Rightarrow 15A + 5B = 21 \leftarrow (2)$$

Multiplying eq. **(2)** by 3 and subtracting from eq. **(1)**

$$\begin{array}{rl} & 55A + 15B = 81 \\ - & 45A + 15B = 63 \\ \hline & 10A \qquad\quad = 18 \end{array} \quad \Rightarrow A = \frac{18}{10} = 1.8$$

Sub. A=1.8 in eq.(2) $\Rightarrow \underset{\substack{\downarrow \\ 15\left(\frac{18}{10}\right)}}{15(1.8)} + 5B = 21 \Rightarrow 5B = 21 \Rightarrow B = -\frac{6}{5} = -1.2$

The least squares line is $f(x) = Ax + B = 1.8x - 1.2$

The root mean square error $E_2(f) = \left[\frac{1}{N}\sum_{k=1}^{N} |f(x_k) - y_k|^2\right]^{\frac{1}{2}}$

x_k	y_k	$f(x_k) = 1.8x_k - 1.2$	$f(x_k) - y_k$	$(f(x_k) - y_k)^2$
1	1	1.8(1)-1.2=0.6	0.6 - 1 = 0.4	0.16
2	2	1.8(2)-1.2=2.4	2.4 – 2 = 0.4	0.16
3	4	1.8(3)-1.2=4.2	4.2 – 4 = 0.2	0.04
4	6	1.8(4)-1.2=6	6 – 6 = 0	0.00
5	8	1.8(5)-1.2=7.8	7.8 – 8 = -0.2	0.04
				0.40

So $E_2(f) = \sqrt{\frac{0.4}{5}} = \sqrt{0.08} = 0.2828$

Matlab

Textbook has program on page 258

function [A B]= lsline(x, y)

⋮

Program uses mean of $x \Rightarrow \bar{x} = \frac{\sum_{k=1}^{N} x_k}{N}$

& mean of $x \Rightarrow \bar{x} = \frac{\sum_{k=1}^{N} x_k}{N}$

For example (1)

x_k	-2	-1	0	1	2
y_k	1	2	3	3	4

```
>> X = [-2 -1 0 1 2];
>> Y = [ 1 2 3 3 4];
>>[A B]=lsline(X,Y)
        A=0.7000
        B=2.6000
```

Textbook page 260 Q#4 $\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$ & $\bar{y} = \frac{1}{N}\sum_{k=1}^{N} y_k$

then $(\bar{x}, \bar{y})$ lies on the line $y = Ax + B \Rightarrow \bar{y} = A\bar{x} + B$

Q#7 $C = \sum_{k=1}^{N}(x_k - \bar{x})^2$ then $A = \frac{1}{C}\left[\sum_{k=1}^{N}(x_k - \bar{x})(y_k - \bar{y})\right]$

once you have A, then $B = \bar{y} - A\bar{x}$

For example (1) → $\bar{x} = \frac{-2-1+0+1+2}{5} = \frac{0}{5} = 0$ & $\bar{y} = \frac{1+2+3+3+4}{5} = \frac{13}{5}$

x_k	y_k	$x_k - \bar{x}$	$y_k - \bar{y}$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})(y_k - \bar{y})$
-2	1	-2	$1 - \frac{13}{5} = -\frac{8}{5}$	4	$\frac{16}{5}$
-1	2	-1	$2 - \frac{13}{5} = -\frac{3}{5}$	1	$\frac{3}{5}$
0	3	0	$3 - \frac{13}{5} = \frac{2}{5}$	0	0
1	3	1	$3 - \frac{13}{5} = \frac{2}{5}$	1	$\frac{2}{5}$
2	4	2	$4 - \frac{13}{5} = \frac{7}{5}$	4	$\frac{14}{5}$
Sum				10	$\frac{38}{5} = 7$

So $A = \frac{7}{10} = 0.7$ & $B = \bar{y} - A\bar{x} = \frac{13}{5} - 0.7(0) = \frac{13}{5} = 2.6$

The least square line is $f(x) = y = 0.7x + 2.6$

By hand calculation, it is easier to solve the normal equations.