Finding the Least-Squares Line

- Given n distinct points: $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$.
- The least-squares line $y = f(x) = Ax + B$ is the line that minimizes the RMSE $E_2(f)$:

$$E_2(f) = \sqrt{\frac{\sum (f(x_i) - y_i)^2}{n}} \Rightarrow n E_2^2 = \sum_{i=1}^{n} (f(x_i) - y_i)^2$$

- E_2 is minimized iff $E(A,B) = \sum_{i=1}^{n} (Ax_i + B - y_i)^2$ is minimized

$$\frac{\partial E}{\partial A} = 2 \sum_{i=1}^{n} (Ax_i + B - y_i)\, x_i = 0 \iff \sum_{i=1}^{n} (Ax_i^2 + Bx_i - y_i x_i) = 0$$

$$A \sum_{i=1}^{n} x_i^2 + B \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \quad \cdots \text{(1)}$$

$$\frac{\partial E}{\partial B} = 2 \sum_{i=1}^{n} (Ax_i + B - y_i) = 0 \iff \sum_{i=1}^{n} (Ax_i + B - y_i) = 0$$

$$A \sum_{i=1}^{n} x_i + nB = \sum_{i=1}^{n} y_i \quad \cdots \text{(2)}$$

- Equations (1) and (2) are called the linear normal equations and used to find the coefficients A and B.

Exp Find the least-squares line for the following data points: (1,2), (3,-1), (2,-1), (0,1), (-1,3)

x	y	xy	x^2
1	2	2	1
3	-1	-3	9
2	-1	-2	4
0	1	0	0
-1	3	-3	1
5	4	-6	15

Normal Equations:

$$A\sum_{i=1}^{5} x_i^2 + B\sum_{i=1}^{5} x_i = \sum_{i=1}^{5} x_i y_i$$

$$\boxed{15A + 5B = -6} \quad (1)$$

$$A\sum_{i=1}^{5} x_i + nB = \sum_{i=1}^{5} y_i$$

$$\boxed{5A + 5B = 4} \quad (2)$$

- $$A = \frac{\begin{vmatrix} -6 & 5 \\ 4 & 5 \end{vmatrix}}{\begin{vmatrix} 15 & 5 \\ 5 & 5 \end{vmatrix}} = \frac{-50}{50} = -1$$

$$B = \frac{\begin{vmatrix} 15 & -6 \\ 5 & 4 \end{vmatrix}}{50} = \frac{90}{50} = 1.8$$

Hence, $y = Ax + B$
$= -x + 1.8$

Exp Find the normal equation for the best fit of the form $y = Ax^m$ where m is known constant.

- $$E(A) = \sum_{i=1}^{n} \left(f(x_i) - y_i\right)^2 = \sum_{i=1}^{n} \left(A x_i^m - y_i\right)^2$$

- $$E'(A) = 2\sum_{i=1}^{n} \left(A x_i^m - y_i\right) x_i^m = 0 \iff$$

$$\sum_{i=1}^{n} A x_i^{2m} - \sum_{i=1}^{m} x_i^m y_i = 0 \iff$$

$$\boxed{A = \frac{\sum_{i=1}^{n} x_i^m y_i}{\sum_{i=1}^{n} x_i^{2m}}}$$

Exp Find the power fits $y = Ax^2$ for the following data. Then find $E_2(f)$.

i	x_i	y_i	x_i^2	$x_i^2 y_i$	x_i^4	$f(x_i)$	$\lvert e_i\rvert^2 = \lvert f(x_i) - y_i\rvert^2$
1	2.0	5.1	4	20.4	16	6.748	2.715904
2	2.3	7.5	5.29	39.675	27.9841	8.92423	2.0284310929
3	2.6	10.6	6.76	71.656	45.6976	11.40412	0.6466089744
4	2.9	14.4	8.41	121.104	70.7281	14.18767	0.0450840289
5	3.2	19.0	10.24	194.56	104.8576	17.27488	2.9760390144
				447.395	265.2674		8.4120671106

- $$A = \frac{\sum_{i=1}^{5} x_i^2 y_i}{\sum_{i=1}^{5} x_i^4} = \frac{447.395}{265.2674} \approx 1.687$$

Hence, $y = f(x) = Ax^2 = 1.687x^2$

- $$E_2(f) = \sqrt{\frac{\sum_{i=1}^{5} |e_i|^2}{5}} = \sqrt{\frac{8.4120671106}{5}} = \sqrt{1.68241342212}$$

$$= 1.2970788034$$

Exp Given the points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Find the normal equations for the best fit of the form $y = f(x) = Ax^2 + Bx + C$

- $$E(A,B,C) = \sum_{i=1}^{n} \left(A x_i^2 + B x_i + C - y_i \right)^2$$

$$\frac{\partial E}{\partial A} = 2 \sum_{i=1}^{n} \left(A x_i^2 + B x_i + C - y_i \right) x_i^2 = 0 \iff$$

$$\boxed{\left(\sum_{i=1}^{n} x_i^4\right) A + \left(\sum_{i=1}^{n} x_i^3\right) B + \left(\sum_{i=1}^{n} x_i^2\right) C = \sum_{i=1}^{n} y_i x_i^2} \quad (1)$$

$$\frac{\partial E}{\partial B} = 2 \sum_{i=1}^{n} \left(A x_i^2 + B x_i + C - y_i \right) x_i = 0 \iff$$

$$\boxed{\left(\sum_{i=1}^{n} x_i^3\right) A + \left(\sum_{i=1}^{n} x_i^2\right) B + \left(\sum_{i=1}^{n} x_i\right) C = \sum_{i=1}^{n} y_i x_i} \quad (2)$$

$$\frac{\partial E}{\partial C} = 2 \sum_{i=1}^{n} \left(A x_i^2 + B x_i + C - y_i \right) = 0 \iff$$

$$\boxed{\left(\sum_{i=1}^{n} x_i^2\right) A + \left(\sum_{i=1}^{n} x_i\right) B + nC = \sum_{i=1}^{n} y_i} \quad (3)$$

- To find A, B, C we solve the three equations above.

Exp Find the least-squares parabola for the four points $(-3,3)$, $(0,1)$, $(2,1)$ and $(4,3)$.

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$	
-3	3	9	-27	81	-9	27	
0	1	0	0	0	0	0	
2	1	4	8	16	2	4	
4	3	16	64	256	12	48	
3	8	29	45	353	5	79	Total

- The least-squares parabola is $y = f(x) = Ax^2 + Bx + C$
- To find A, B, C we use equations ①, ②, ③ in page 112:

$$\left.\begin{aligned} 353A + 45B + 29C &= 79 \\ 45A + 29B + 3C &= 5 \\ 29A + 3B + 4C &= 8 \end{aligned}\right\} \Rightarrow \quad A = \frac{585}{3278} = 0.178462, \quad B = -0.192495, \quad C = 0.850519$$

Hence, $y = f(x) = 0.178462\, x^2 - 0.192495\, x + 0.850519$

Exp Find the best fit of the form $y = A \sin(\pi x)$

$$E(A) = \sum_{i=1}^{n} \left(A \sin(\pi x_i) - y_i \right)^2$$

$$E'(A) = 2 \sum_{i=1}^{n} \left(A \sin(\pi x_i) - y_i \right) \sin(\pi x_i) = 0$$

$$A = \frac{\sum_{i=1}^{n} y_i \sin(\pi x_i)}{\sum_{i=1}^{n} \sin^2(\pi x_i)}$$