Practice Problems for Curve Fitting - Solution

A. Linear Regression

1. **Use least squares regression to fit a straight line to the data given in Table 1. Along with the slope and intercept, compute the standard error of the estimate and the coefficient of determinant.**

Table 1

x	0	2	4	6	9	11
y	5	6	7	6	9	8

2. **An investigator has reported the data tabulated in Table 2 for an experiment to determine the growth rate of bacteria k (per d) as a function of oxygen concentration (mg/L). It is known that such data can be modeled by the following equation:**

$$k = \frac{k_{\max} c^2}{c_s + c^2}$$

where c_s and k_{max} are parameters. Use a transformation to linearize this equation. Then use linear regression to estimate c_s and k_{max} and predict the growth rate at c=2 mg/L.

Table 2

C	0.5	0.8	1.5	2.5	4
K	1.1	2.4	5.3	7.6	8.9

B. General Linear Least Squares and nonlinear regression

A physics process can be described with the equation $y = f(x) = \frac{a_0}{x} + \frac{a_1}{x^2}$. The measured values of (x, y) are listed in the following table:

Table 3

x	1	2	3	4
y	3	0.9	0.6	0.4

Use direct nonlinear regression method to determine a_0 and a_1.

C. Polynomial Interpolation

Assume that the tabulated data in Table 3 are precise; apply the 3rd order Lagrange Polynomial to approximate $f(2.5)$ with precision to five decimal spaces.

D. Spline interpolation

Fit the data in Table 4 with a quadratic with natural end conditions.

Table 4

x	**1**	**1.5**	**2**	**2.5**
y	55	22	13	10

Practice Problems for Curve Fitting - Solution

A. Linear Regression

1. Use least squares regression to fit a straight line to the data given in Table 1. Along with the slope and intercept, compute the standard error of the estimate and the coefficient of determinant.

Table 3

x	0	2	4	6	9	11
y	5	6	7	6	9	8

Solution:

The straight line that results in the least sum of squares of the residuals between the measured y and the calculated y will be:

$$y = a_1 x + a_0$$

where $a_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$ and $a_0 = \overline{y} - a_1\overline{x}$

For the data in table 1 we have:

$n = 6$, $\sum x_i = 32$, $\sum y_i = 41$, $\sum x_i y_i = 245$, $\sum x_i^2 = 258$, $\overline{x} = 5.33$, $\overline{y} = 6.83$.

Substituting in the above relations:

$$a_1 = \frac{6\cdot 245 - 32\cdot 41}{6\cdot 258 - 32^2} = 0.30152 \text{ and } a_0 = 6.83 - 0.30152\cdot 5.33 = 5.2258$$

Therefore, the equation we are looking for is: $y = 0.3015x + 5.2258$.

The standard error of the estimate, $S_{Y/X}$, is given by the following equation:

$$S_{Y/X} = \sqrt{\frac{S_r}{n-2}}$$

where S_r is defined as

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)^2$$

In this case, S_r is computed to be $S_r = 2.8931$.

The coefficient of determinant is r^2 is the normalized decrease in the error due to describing the data in terms of a straight line rather than an average value and it is given by the following equation:

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

where S_t is the total sum of squares around the mean for y:

$$S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

For the given data, we have:

$$S_t = 10.833 \text{ and } r^2 = \frac{10.833 - 2.893}{10.833} = 0.7329$$

2. **An investigator has reported the data tabulated in Table 2 for an experiment to determine the growth rate of bacteria k (per d) as a function of oxygen concentration (mg/L). It is known that such data can be modeled by the following equation:**

$$k = \frac{k_{\max} c^2}{c_s + c^2}$$

where c_s and k_{max} are parameters. Use a transformation to linearize this equation. Then use linear regression to estimate c_s and k_{max} and predict the growth rate at c=2 mg/L.

Table 4

C	0.5	0.8	1.5	2.5	4
K	1.1	2.4	5.3	7.6	8.9

Solution:

The given equation for the growth rate of bacteria can be linearized in the following way:

$$k = \frac{k_{\max} c^2}{c_s + c^2} \Rightarrow \frac{1}{k} = \frac{c_s}{k_{\max}} \cdot \frac{1}{c^2} + \frac{1}{k_{\max}}$$

which is of the form $y = a_1 x + a_0$

with $y = 1/k$, $x = 1/c^2$, $a_0 = 1/k_{\max}$ and $a_1 = c_s/k_{\max}$.

As a result if we remake the data in table 2 accordingly, we can use linear regression to estimate c_s and k_{max}.

$1/c^2$	4	1.563	0.444	0.160	0.063
$1/k$	0.909	0.417	0.189	0.132	0.112

As in exercise 1, we need to calculate the following quantities for this data:

$n = 5$, $\sum x_i = 6.299$, $\sum y_i = 1.758$, $\sum x_i y_i = 4.399$, $\sum x_i^2 = 18.668$,
$\bar{x} = 2.076$, $\bar{y} = 0.586$.

Then we can calculate α_0 and α_1:

$$a_1 = \frac{5 \cdot 4.399 - 6.299 \cdot 1.758}{5 \cdot 18.668 - 6.299^2} = 0.202, \quad a_0 = 0.586 - 0.202 \cdot 2.076 = 0.099$$

Now that we know α_0 and α_1, we can calculate c_s and k_{max}:

$$a_0 = 1/k_{\max} \Rightarrow k_{\max} = 1/a_0 = 1/0.099 = 10.1 \text{ d}^{-1}$$

$$a_1 = c_s/k_{\max} \Rightarrow c_s = a_1 \cdot k_{\max} = 0.202 \cdot 10.1 = 2.04 \,\text{mg/L}$$

and the expression for k is:

$$k = \frac{10.1 \cdot c^2}{2.04 + c^2} \overset{c=2mg/L}{\Rightarrow} k = 6.689 \ d^{-1}$$

B. General Linear Least Squares and nonlinear regression

A physics process can be described with the equation $y = f(x) = \frac{a_0}{x} + \frac{a_1}{x^2}$**. The measured values of** (x, y) **are listed in the following table:**

Table 3

x	1	2	3	4
y	3	0.9	0.6	0.4

Use direct nonlinear regression method to determine a_0 **and** a_1**.**

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

$$\begin{cases} \dfrac{\partial S_r}{\partial a_0} = \sum 2(y_i - \dfrac{a_0}{x_i} - \dfrac{a_1}{x_i^2})(-\dfrac{1}{x_i}) = 0 \\ \dfrac{\partial S_r}{\partial a_1} = \sum 2(y_i - \dfrac{a_0}{x_i} - \dfrac{a_1}{x_i^2})(-\dfrac{1}{x_i^2}) = 0 \end{cases} \Rightarrow \begin{cases} \sum (y_i - \dfrac{a_0}{x_i} - \dfrac{a_1}{x_i^2})(-\dfrac{1}{x_i}) = 0 \\ \sum (y_i - \dfrac{a_0}{x_i} - \dfrac{a_1}{x_i^2})(-\dfrac{1}{x_i^2}) = 0 \end{cases} \Rightarrow$$

$$\begin{cases} \sum (y_i \dfrac{1}{x_i} - \dfrac{a_0}{x_i^2} - \dfrac{a_1}{x_i^3}) = 0 \\ \sum (y_i \dfrac{1}{x_i^2} - \dfrac{a_0}{x_i^3} - \dfrac{a_1}{x_i^4}) = 0 \end{cases} \Rightarrow \begin{cases} a_0 \sum \dfrac{1}{x_i^2} + a_1 \sum \dfrac{1}{x_i^3} = \sum \dfrac{y_i}{x_i} \\ a_0 \sum \dfrac{1}{x_i^3} + a_1 \sum \dfrac{1}{x_i^4} = \sum \dfrac{y_i}{x_i^2} \end{cases}$$

i	x_i	y_i	$1/x_i^2$	$1/x_i^3$	$1/x_i^4$	y_i/x_i	y_i/x_i^2
1	1	3	1.0000	1.0000	1.0000	3.0000	3.0000
2	2	0.9	0.2500	0.1250	0.0625	0.4500	0.2250
3	3	0.6	0.1111	0.0370	0.0123	0.2000	0.0667
4	4	0.4	0.0625	0.0156	0.0039	0.1000	0.0250
$\sum$			1.4236	1.1777	1.0788	3.7500	3.3167

The system to solve is then:

$$\begin{cases} 1.4236a_0 + 1.1777a_1 = 3.7500 \\ 1.1777a_0 + 1.0778a_1 = 3.3167 \end{cases} \Rightarrow \begin{cases} a_0 = 0.9206 \\ a_1 = 2.0714 \end{cases}$$

C. Polynomial Interpolation

Assume that the tabulated data in Table 3 are precise; apply the 3rd order Lagrange Polynomial to approximate $f(2.5)$ with precision to five decimal spaces.

Solution:

The third order Lagrange polynomial for the four data points given in table 3 will be:

$$y = \sum_{i=1}^{4} y_i L_i(x),\ L_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{4} \frac{x - x_j}{x_i - x_j}$$

where:

- for i=1

$$L_1(x)=\frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}=\frac{(x-2)(x-3)(x-4)}{-6}$$

- for i=2

$$L_2(x)=\frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}=\frac{(x-1)(x-3)(x-4)}{2}$$

- for i=3

$$L_3(x)=\frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}=\frac{(x-1)(x-2)(x-4)}{-2}$$

- and for i=4

$$L_4(x)=\frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}=\frac{(x-1)(x-2)(x-3)}{6}$$

Changing the notation of the data points to make it easier to substitute, gives:

Substituting gives:

$$f_3(x)=\frac{(x-2)(x-3)(x-4)}{-6}\cdot 3+\frac{(x-1)(x-3)(x-4)}{2}\cdot 0.9+$$
$$+\frac{(x-1)(x-2)(x-4)}{-2}\cdot 0.6+\frac{(x-1)(x-2)(x-3)}{6}\cdot 0.4\Rightarrow$$

$$f_3(x)=-0.5(x-2)(x-3)(x-4)+0.45(x-1)(x-3)(x-4)-$$
$$-0.3(x-1)(x-2)(x-4)+0.067(x-1)(x-2)(x-3)$$

To find f(2.5) we substitute x=2.5 in the above equation:

$$f_3(x)=-0.5(2.5-2)(2.5-3)(2.5-4)+0.45(2.5-1)(2.5-3)(2.5-4)-$$
$$-0.3(2.5-1)(2.5-2)(2.5-4)+0.067(2.5-1)(2.5-2)(2.5-3)=0.63125$$

D. Spline interpolation

Fit the data in Table 4 with a quadratic with natural end conditions.

Table 4

x	1	1.5	2	2.5
y	55	22	13	10

Solution:

For quadratic splines, we want to derive a second order polynomial for each interval which will fit the data in the interval. The polynomial is of the following form:

$$f_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2$$

Natural condition implies $c_1 = 0,\ \ b_1 = (y_2 - y_1)/(x_2 - x_1)$.

Use the formula:

$$\begin{cases} a_{i+1} = y_{i+1} \\ b_{i+1} = b_i + 2c_i(x_{i+1} - x_i) \\ c_{i+1} = \dfrac{y_{i+2} - y_{i+1}}{(x_{i+2} - x_{i+1})^2} - \dfrac{b_{i+1}}{(x_{i+2} - x_{i+1})} \end{cases} \Rightarrow \begin{cases} a_1 = 55 \\ b_1 = -66; \\ c_1 = 0 \end{cases} \begin{cases} a_2 = 22 \\ b_2 = -66; \\ c_2 = 96 \end{cases} \begin{cases} a_3 = 13 \\ b_3 = 30 \\ c_3 = -72 \end{cases}$$

The splines are plotted in the graph below.