Unit 2

Interpolation & Regression

Interpolation Vs Extrapolation

The method or technique to estimate the unknown result from a given set of observation within a known range is called interpolation.

In many scientific & engineering experiments, we observe the value of dependent variable for the known discrete values of independent variable. The value of dependent variable within the known range of independent variable can be estimated by using the numerical method known as interpolation. The function which is used for interpolation is called as interpolation polynomial.

Interpolation	Extrapolation
It is the method of estimating the unknown value between the known range	It is the method of estimating the unknown value beyond the known range of data set.
It is useful to estimate the missing past value	It is primarily used in forecasting.
estimated result is more likely to be correct.	Estimated result is only a probability.
Eg: To estimate the population at a year between two known years, we use interpolation.	Eg: To estimate the population for the future from the known data set, we use extrapolation.

Lagrange's Interpolation

Let $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the $n+1$ observation points. A polynomial equation that satisfies all these points and also helps to estimate the values corresponding to other unknown points is known as interpolation polynomial.

Let the interpolation polynomial of degree n be,

$$f(x) = A_0 (x-x_1)(x-x_2)\cdots(x-x_n) + A_1 (x-x_0)(x-x_2)\cdots(x-x_n) + \cdots + A_n (x-x_0)(x-x_1)\cdots(x-x_{n-1}) \quad \text{——— (i)}$$

For $x = x_0$, equation (i) gives

$$A_0 = \frac{y_0}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)}$$

Similarly for $x = x_1, x = x_2, \ldots, x = x_n$, we get

$$A_1 = \frac{y_1}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}$$

$$\cdots\cdots$$

$$A_n = \frac{y_n}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}$$

With the values of $A_0, A_1, \ldots, A_n$, eqn (i) becomes,

$$f(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)} y_0$$

$$+ \frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} y_1$$

$$+ \cdots$$

$$+ \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})} y_n$$

Which is the Lagrange's interpolation formula. It is used to find the interpolation polynomial for the observation points that are not necessarily equally spaced.

CSIT-067

Q. Derive the lagrange interpolation equation for given observation and evaluate $f(x)$ at $x = 1$. [4+4]

x	-1	-2	2	4
$f(x)$	-1	-9	11	69

→ Derivation of Lagrange's interpolation polynomial is as above. Second/numerical part is here.

Here, $x = 1$

$x_0 = -1$ $\quad x_1 = -2$ $\quad x_2 = 2$ $\quad x_3 = 4$

$y_0 = -1$ $\quad y_1 = -9$ $\quad y_2 = 11$ $\quad y_3 = 69$

Using Lagrange's interpolation formula,

$$f(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} y_0$$

$$+ \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} y_1$$

$$+ \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} y_2$$

$$+ \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} y_3$$

or $$f(1) = \frac{(1+2)(1-2)(1-4)}{(-1+2)(-1-2)(-1-4)} \times (-1)$$

$$+ \frac{(1+1)(1-2)(1-4)}{(-2+1)(-2-2)(-2-4)} \times (-9)$$

$$+ \frac{(1+1)(1+2)(1-4)}{(2+1)(2+2)(2-4)} \times 11$$

$$+ \frac{(1+1)(1+2)(1-2)}{(4+1)(4+2)(4-2)} \times 69$$

$$= (-0.6) + (2.25) + (8.25) + (6.5)$$

$$= 16.4$$

Q. The values of e^x are given in table below.

x	0	1	2	3
e^x	1	2.7183	7.3891	20.0855

Determine the value of $e^{1.2}$ by using second order polynomial interpolation using Lagrange's method.

→ Hint: Solving procedure is as previous but one thing important here is that, question says second order polynomial. Hence we use $n=2$ in the standard lagrange's formula. The formula to be used is:

$$f(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} y_2$$

Where,

$x = 1.2$

$x_0 = 0$ $\quad$ $y_0 = 1$

$x_1 = 1$ $\quad$ $y_1 = 2.7183$

$x_2 = 2$ $\quad$ $y_2 = 7.3891$

(x_3 & y_3 are not used).

[Ans: 3.41626]

Q. The upward velocity of a rocket is given in the following table.

time(t)	0	10	15	20	22.5	30
velocity (V)	0	227.04	362.78	517.35	602.97	901.67

Estimate the velocity at $t=16$, using first order Lagrange's polynomial.

→ Hint: Take $n = 1$. Ans: 393.69.

CSIT-071 Old

Q. Define interpolation. Find the Lagrange interpolation polynomial to fit the following data. Estimate the value of $e^{1.9}$. [1 + 6 + 1]

i	0	1	2	3
x_i	0	1	2	3
e^{x_i}	0	1.7183	6.3891	19.0855

CSIT-074 Old

Q. Define interpolation. Find the Lagrange interpolation polynomial to fit the following data.

i	0	1	2	3
x_i	0	1	2	3
$e^{x_i}-1$	0	1.7183	6.3891	19.0855

Use the polynomial to estimate the value of $e^{1.5}$.

Page No.
Date: / /

Newton's Divided Difference Interpolation

This method is used to estimate the interpolation polynomial where the given observation points are not necessarily equally spaced.

Let $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the given $n+1$ observation points. The Newton's interpolation polynomial by using divided diff. is given by,

$$f(x) = y_0 + (x-x_0) f[x_0, x_1] + (x-x_0)(x-x_1) f[x_0, x_1, x_2]$$
$$+ \cdots$$
$$+ (x-x_0)(x-x_1)(x-x_2)\cdots(x-x_{n-1}) f[x_0, x_1, x_2, \ldots, x_n]$$

Where,

$$f[x_0, x_1] = \frac{y_1 - y_0}{x_1 - x_0} \quad (1^{st} \text{ divided diff. or } \Delta y_0)$$

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} \quad (2^{nd} \text{ d.d. or } \Delta^2 y_0)$$

$$\cdots \quad \vdots \quad \vdots$$

$$f[x_0, x_1, x_2, \ldots, x_n] = \frac{f[x_1, x_2, \ldots, x_n] - f[x_0, x_1, \ldots, x_{n-1}]}{x_n - x_0}$$

$$(n^{th} \text{ d.d. or } \Delta^n y_0)$$

Divided Difference Table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
x_0	y_0			
		$f[x_0, x_1]$		
x_1	y_1		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$
x_2	y_2		$f[x_1, x_2, x_3]$	
		$f[x_2, x_3]$		
x_3	y_3			

The higher order divided differences can be calculated recursively using the lower order divided differences.

Q. From the given observation points, derive an equation for interpolation polynomial using the Newton's divided difference method. Also estimate the value of $f(1.5)$.

x	-1	1	2	3
$f(x)$	-21	15	12	3

$\rightarrow$ Here,

$x_0 = -1 \quad x_1 = 1 \quad x_2 = 2 \quad x_3 = 3$

$y_0 = -21 \quad y_1 = 15 \quad y_2 = 12 \quad y_3 = 3$

Constructing the divided difference table,

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
$x_0 = -1$	-21	18	-7	1
$x_1 = 1$	15	-3	-3	
$x_2 = 2$	12	-9		
$x_3 = 3$	3			

Newton's interpolation polynomial using divided diff. is given by,

$$f(x) = y_0 + (x-x_0) f[x_0,x_1] + (x-x_0)(x-x_1) f[x_0,x_1,x_2] + (x-x_0)(x-x_1)(x-x_2) f[x_0,x_1,x_2,x_3] \quad \text{--- (i)}$$

Here,

$$f[x_0,x_1] = (y_1-y_0)/(x_1-x_0) = 18$$
$$f[x_0,x_1,x_2] = \cancel{18}\ -7$$
$$f[x_0,x_1,x_2,x_3] = 1$$

Eqn (i) becomes,

$$f(x) = -21 + 18(x+1) + (-7)(x+1)(x-1) + 1.(x+1)(x-1)(x-2)$$
$$\therefore f(x) = x^3 - 9x^2 + 17x + 6$$

which is the required interpolation polynomial.

For $x = 1.5$,

$$f(1.5) = (1.5)^3 - 9\times(1.5)^2 + 17\times(1.5) + 6$$
$$= 14.625$$

Page No.

Date: / /

Q. Given the following data points, create the table of divided differences. Use the table to estimate $f(1.8)$ using second and 3rd order polynomials.

x	1	2	3	4
$f(x)$	0	7	26	63

→ Hint:

Second order polynomial

$$f(x) = y_0 + (x-x_0) f[x_0,x_1] + (x-x_0)(x-x_1) f[x_0,x_1,x_2]$$

find $f(1.8) = 4.64$

Third order polynomial

$$f(x) = y_0 + (x-x_0) f[x_0,x_1] + (x-x_0)(x-x_1) f[x_0,x_1,x_2] + (x-x_0)(x-x_1)(x-x_2) f[x_0,x_1,x_2,x_3]$$

find $f(1.8) = 4.832$

Q. Using Newton's divided difference method; estimate $f(8)$ for the given observation points.

x	4	5	7	10	11	13
$f(x)$	48	100	294	900	1210	2028

[Ans: 448]

Newton's Forward Difference Interpolation Method

This method is used to estimate the interpolation polynomial where the given observation points are equally spaced and the value must be estimated at a point lying in the first half of the observation points.

Let $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the given $n+1$ observation points. The Newton's forward difference interpolation polynomial is given by,

$$f(x) = y_0 + p\,\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 + \cdots$$

Where,

$$p = \frac{x - x_0}{h} \Rightarrow x = x_0 + ph$$

h = Interval size.

$\Delta y_0 = y_1 - y_0$ (First forward difference)
$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$ (second forward difference)
$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$ (Third forward difference).

To estimate the interpolation polynomials, we need to evaluate the different order forward differences denoted by $\Delta y_0, \Delta^2 y_0, \ldots$ etc. For this, we can construct a forward difference table.

Forward Difference Table

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
x_0	y_0			
		$\Delta y_0 = y_1 - y_0$		
x_1	y_1		$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$	
		$\Delta y_1 = y_2 - y_1$		$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$
x_2	y_2		$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$	
		$\Delta y_2 = y_3 - y_2$		
x_3	y_3			

Q. $f(x)$ is known at the following data points.

x	0	1	2	3	4
$f(x)$	1	7	23	55	109

Estimate $f(0.5)$ and $f(1.5)$ using Newton's forward difference formula.

→ Constructing the forward difference table,

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
$x_0 = 0$	$y_0 = 1$				
		$\Delta y_0 = 6$			
$x_1 = 1$	$y_1 = 7$		$\Delta^2 y_0 = 10$		
		$\Delta y_1 = 16$		$\Delta^3 y_0 = 6$	
$x_2 = 2$	$y_2 = 23$		$\Delta^2 y_1 = 16$		$\Delta^4 y_0 = 0$
		$\Delta y_2 = 32$		$\Delta^3 y_1 = 6$	
$x_3 = 3$	$y_3 = 55$		$\Delta^2 y_2 = 22$		
		$\Delta y_3 = 54$			
$x_4 = 4$	$y_4 = 109$				

The Newton's forward difference interpolation formula can be written as,

$$f(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 \quad \text{—— (i)}$$

Where, $p = \frac{x - x_0}{h}$

or, $p = \frac{x-0}{1}$

$\therefore p = x$

Hence, eqn (i) gives

$$f(x) = y_0 + x\Delta y_0 + \frac{x(x-1)}{2!}\Delta^2 y_0 + \frac{x(x-1)(x-2)}{3!}\Delta^3 y_0$$

$$= 1 + 6x + 5x^2 - 5x + x^3 - 3x^2 + 2x$$

$$= x^3 + 2x^2 + 3x + 1$$

$\therefore f(x) = x^3 + 2x^2 + 3x + 1$ is the required interpolation polynomial.

Now, $f(0.5) = 0.5^3 + 2 \times 0.5^2 + 3 \times 0.5 + 1$

$= 3.125$

$f(1.5) = 1.5^3 + 2 \times 1.5^2 + 3 \times 1.5 + 1$

$= 3.375 + 4.5 + 4.5 + 1$

$= 13.375$

Note that, the interpolation polynomial with n+1 given points will have at most degree n.

Q. From the table, estimate the number of students who scorred marks between 40 and 45.

marks	30-40	40-50	50-60	60-70	70-80
No. of std	31	42	51	35	31

→ Hint, rearrange the table for discrete points as

marks below x	40	50	60	70	80
student $f(x)$	31	73	124	159	190

Now find $f(x)$ using Newton's forward difference formula, using forward difference table.

$x = 45$ (as asked by question)

$x_0 = 40$

$h = 10$

$p = \dfrac{45-40}{10} = 0.5$

Use formula

$$f(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \dots$$

Ans: $f(45) = 47.87 \simeq 48$.

Q. Find the population increase between 1946 to 1948. (Use forward difference interpolation).

year	1911	1921	1931	1941	1951	1961
population(K)						

→ Hint: find population at 1946 & then 1948. The diff. in population betn 1948 & 1946 is the answer!

Ans: 2.530×1000

Newton's Backward Difference Interpolation Method:

This method is used to estimate the interpolation polynomial where the given observation points are equally spaced and the value must be estimated at a point lying in the second half of the observation points.

Let $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be the $n+1$ given observation points. The Newton's backward difference interpolation polynomial is given by,

$$f(x) = y_n + p\nabla y_n + \frac{p(p+1)}{2!}\nabla^2 y_n + \frac{p(p+1)(p+2)}{3!}\nabla^3 y_n$$

$$+ \cdots$$

Where,

$$p = \frac{x - x_n}{h} \Rightarrow x = x_n + ph$$

h = Interval size

$\nabla y_n = y_n - y_{n-1}$ (first backward difference)
$\nabla^2 y_n = \nabla y_n - \nabla y_{n-1}$ (second backward difference)
$\nabla^3 y_n = \nabla^2 y_n - \nabla^2 y_{n-1}$ (Third backward difference)

To estimate the interpolation polynomials, we need to evaluate the backward differences of different order, denoted by $\nabla y_n, \nabla^2 y_n, \nabla^3 y_n$ etc. For this, we can construct a backward difference table.

Backward Difference Table

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$
x_0	y_0			
		$\nabla y_1 = y_1 - y_0$		
x_1	y_1		$\nabla^2 y_2 = \nabla y_2 - \nabla y_1$	
		$\nabla y_2 = y_2 - y_1$		$\nabla^3 y_3 = \nabla^2 y_3 - \nabla^2 y_2$
x_2	y_2		$\nabla^2 y_3 = \nabla y_3 - \nabla y_2$	
		$\nabla y_3 = y_3 - y_2$		
x_3	y_3			

Q. Find $f(0.15)$ using Newton's backward interpolation.

x	0.1	0.2	0.3	0.4	0.5
$f(x)$	0.09983	0.19867	0.29552	0.38942	0.47943

→ Constructing backward difference table,

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
0.1	0.09983				
		0.09884			
0.2	0.19867		-0.00199		
		0.09685		-0.00156	
0.3	0.29552		-0.00355		0.00121
		0.0939		-0.00035	
0.4	0.38942		-0.0039		
		0.09001			
0.5	0.47943				

$\therefore\ x_n = 0.5$

$y_n = 0.47943$ $\qquad \nabla^2 y_n = -0.0039$

$\nabla y_n = 0.09001$ $\qquad \nabla^3 y_n = -0.00035$

$\nabla^4 y_n = 0.00121$

from the question,

$h = 0.1$, $x = 0.15$, $x_n = 0.5$, $y_n = 0.47943$

So, $p = \frac{x - x_n}{h}$

$= \frac{0.15 - 0.5}{0.1}$

$= -3.5$

Using backward difference formula,

$$f(x) = y_n + p\nabla y_n + \frac{p(p+1)}{2!}\nabla^2 y_n + \frac{p(p+1)(p+2)}{3!}\nabla^3 y_n + \frac{p(p+1)(p+2)(p+3)}{3!}\nabla^4 y_n$$

or, $$f(0.15) = 0.47943 + (-3.5)\times 0.09 + \frac{(-3.5)(-3.5+1)}{2}.(-0.0039) + \frac{(-3.5)(-3.5+1)(-3.5+2)}{6}.(-0.00035) + \frac{(-3.5)(-3.5+1)(-3.5+2)(-3.5+3)}{24}.(0.00121)$$

$= 0.14847$

Cubic Spline Interpolation

In the interpolation methods such as Lagrange interpolation, Newton's interpolation etc. we construct a single interpolation polynomial for the given entire data points.

In the spline interpolation method, we estimate a separate interpolation function for each pair of data points. If we estimate a straight line (linear interpolation) between each pair of data points, the method is known as linear spline interpolation.

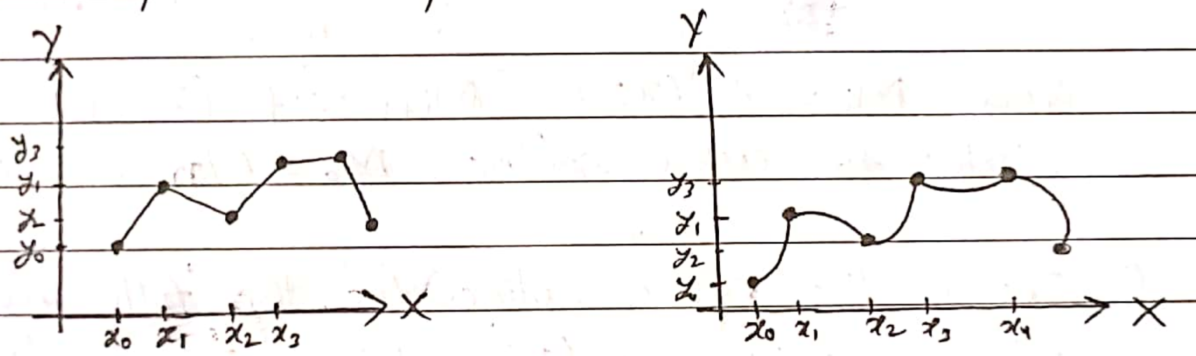


fig: Linear Spline Interpolation   fig: Cubic spline Interpolation

For $n+1$ given data points $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, we have n intervals and so there are n splines. If the degree of interpolation polynomial for each interval is 3, the method is called as cubic spline interpolation method.

Let $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ be the $n+1$ different observation points. Let $f(x)$ be the cubic polynomial in each of the subintervals, such that,

(i) $f(x)$ is linear outside the interval (x_0, x_n).
(ii) $f(x)$ is a cubic polynomial in each subinterval.
(iii) $f'(x)$ & $f''(x)$ are continuous at each point.

With the above assumptions, the cubic splines for the given data points for equally spaced interval is given by:

$$f(x) = \frac{(x_{i+1}-x)^3}{6h} M_i + \frac{(x-x_i)^3}{6h} M_{i+1}$$
$$+ \frac{(x_{i+1}-x)}{h}\left(y_i - \frac{h^2}{6} M_i\right)$$
$$+ \frac{(x-x_i)}{h}\left(y_{i+1} - \frac{h^2}{6} M_{i+1}\right)$$

Where,

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{h^2}(y_{i-1} - 2y_i + y_{i+1}) \text{ for } i = 1, 2, \ldots, n-1$$

And $M_i = f''(x_i)$, $M_{i+1} = f''(x_{i+1})$

(Note: for cubic spline, $M_0 = M_n = 0$ always).

Q. Obtain the cubic splines for the following data points, also evaluate $f(1.5)$ using cubic spline method.

x	0	1	2	3
y	2	-6	-8	2

→ Here, $h = 1$, $n = 3$

$x_0 = 0$ $\quad x_1 = 1$ $\quad x_2 = 2$ $\quad x_3 = 3$

$y_0 = 2$ $\quad y_1 = -6$ $\quad y_2 = -8$ $\quad y_3 = 2$

Cubic splines for the interval $x_i \le x \le x_{i+1}$ is given by:

$$f(x) = \frac{(x_{i+1}-x)^3}{6h} M_i + \frac{(x-x_i)^3}{6h} M_{i+1}$$
$$+ \frac{(x_{i+1}-x)}{h}\left(y_i - \frac{h^2}{6} M_i\right) + \frac{(x-x_i)}{h}\left(y_{i+1} - \frac{h^2}{6} M_{i+1}\right) \quad \text{(i)}$$

Where,

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{h^2}(y_{i-1} - 2y_i + y_{i+1}), \ i = 1,2 \quad \text{——— (ii)}$$

first we find the values of M using eqn (ii)

for i = 1, eqn (ii) becomes

$$M_0 + 4M_1 + M_2 = \frac{6}{1^2}(y_0 - 2y_1 + y_2)$$

or, $M_0 + 4M_1 + M_2 = 6(2 - 2\times(-6) + (-8))$

$\therefore \cancel{M_0}^{0} + 4M_1 + M_2 = 36$ ——— (iii)

for i = 2, eqn (ii) becomes

$$M_1 + 4M_2 + \cancel{M_3}^{0} = \frac{6}{1^2}(y_1 - 2y_2 + y_3)$$

or, $M_1 + 4M_2 = 6(-6 - 2\times(-8) + 2)$

or, $M_1 + 4M_2 = 72$ ——— (iv)

Solving (iii) and (iv), we get $M_1 = 4.8$, $M_2 = 16.8$

Now, we can find the cubic splines for each interval using eqn (i).

For first sub-interval (betn $x=0$ & $x=1$), let $i=0$ in eqn (i).

$$f(x) = \frac{1}{6}(1-x)^3 . \cancel{M_0}^{0} + \frac{1}{6}(x-0)^3(4.8)$$

$$+ (1-x)(2-0) + x\left(-6 - \frac{4.8}{6}\right)$$

$\therefore f(x) = 0.8x^3 - 8.8x + 2, \quad 0 \le x \le 1$

For second sub-interval, taking $i=1$ in eqn (i) we get,

$$f(x) = \frac{1}{6}(2-x)^3(4.8) + \frac{1}{6}(x-1)^3(16.8) + (2-x)\left(-6-\frac{4.8}{6}\right)$$

$$+ (x-1)\left(-8-\frac{16.8}{6}\right)$$

$$\therefore f(x) = 2x^3 - 5.84x^2 + 1.68x + 0.8 \quad , \quad 1 \le x \le 2.$$

For third sub-interval, taking $i=2$ in eqn (i), we get;

$$f(x) = \frac{1}{6}(3-x)^3 \cdot (16.8) + \frac{1}{6}(x-2)^3 \cdot (0) + (3-x)\left(-8-\frac{16.8}{6}\right)$$

$$+ (x-2)(2-0)$$

$$\therefore f(x) = -0.8x^3 + 2.64x^2 + 9.68x - 14.8 \quad , \quad 2 \le x \le 3.$$

To find the value of $f(1.5)$, we can use the cubic spline for the interval $1 \le x \le 2$.

$$f(x) = 2x^3 - 5.84x^2 + 1.68x + 0.8 \quad , \quad 1 \le x \le 2.$$

$$f(1.5) = 2 \times (1.5)^3 - 5.84 \times (1.5)^2 + 1.68 \times 1.5 + 0.8$$
$$= -3.07$$

The required cubic spline is:

$$f(x) = \begin{cases} 0.8x^3 - 8.8x + 2 & , \quad 0 \le x \le 1 \\ 2x^3 - 5.84x^2 + 1.68x + 0.8 & , \quad 1 \le x \le 2 \\ -0.8x^3 + 2.64x^2 + 9.68x - 14.8 & , \quad 2 \le x \le 3. \end{cases}$$

Q. Find the cubic spline interpolation for the following data set.

x	1	2	3	4	5
$f(x)$	1	0	1	0	1

→ Here, $h = 1$, $n = 4$

$x_0 = 1$ $x_1 = 2$ $x_2 = 3$ $x_3 = 4$ $x_4 = 5$

$y_0 = 1$ $y_1 = 0$ $y_2 = 1$ $y_3 = 0$ $y_4 = 1$

Cubic spline for the interval $x_i \le x \le x_{i+1}$ is given by the polynomial,

$$f(x) = \frac{(x_{i+1} - x)^3}{6h} M_i + \frac{(x - x_i)^3}{6h} M_{i+1} + \frac{(x_{i+1} - x)}{h}\left(y_i - \frac{h^2}{6} M_i\right) + \frac{(x - x_i)}{h}\left(y_{i+1} - \frac{h^2}{6} M_{i+1}\right) \quad \text{— (i)}$$

Where,

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{h^2}\left(y_{i-1} - 2y_i + y_{i+1}\right), \; i = 1,2,3 \quad \text{— (ii)}$$

First we find the value of M using eqn (ii).

for $i = 1$, eqn (ii) gives

$$\cancel{M_0}^{0} + 4M_1 + M_2 = 6(y_0 - 2y_1 + y_2)$$

or, $4M_1 + M_2 = 6(1 - 2\times 0 + 1)$

$\therefore$ $4M_1 + M_2 = 12$ —— (iii)

for $i=2$, eq^n (ii) becomes

$$M_1 + 4M_2 + M_3 = 6(y_1 - 2y_2 + y_3)$$

or, $M_1 + 4M_2 + M_3 = 6(0 - 2\times1 + 0)$

$\therefore M_1 + 4M_2 + M_3 = -12$ ——————— (iv)

for $i=3$, eq^n (ii) becomes

$$M_2 + 4M_3 + \cancel{M_4}^{0} = 6(y_2 - 2y_3 + y_4)$$

or, $M_2 + 4M_3 = 6(1 - 2\times0 + 1)$

$\therefore M_2 + 4M_3 = 12$ ——————— (v)

Solving eq^n (iii), (iv) & (v), we get

$M_1 = 30/7$

$M_2 = -36/7$

$M_3 = 30/7$

Now, we can find the cubic splines for each interval using eq^n (i).

for 1st interval $(1 \le x \le 2)$, put $i=0$ in eq^n (i).

$$f(x) = \frac{1}{6}(2-x)^3 . \cancel{M_0}^{0} + \frac{1}{6}(x-1)^3 . \frac{30}{7}$$

$$+ (2-x)\left(1 - \frac{1}{6}\cancel{M_0}^{0}\right) + (x-1)\left(0 - \frac{1}{6} . \frac{30}{7}\right)$$

$$= \frac{1}{6}(x-1)^3 . \frac{30}{7} + (2-x) + (x-1) . \left(\frac{-30}{6\times7}\right)$$

$$\therefore f(x) = \frac{5}{7}(x-1)^3 + (2-x) - \frac{5}{7}(x-1) \quad , \quad 1 \le x \le 2$$

for second interval $(2 \le x \le 3)$, put $i = 1$ in eq^n(i)
we get,

$$f(x) = \frac{5}{7}(3-x)^3 - \frac{6}{7}(x-2)^3 - \frac{5}{7}(3-x) + \frac{13}{7}(x-2), \quad (2 \le x \le 3)$$

For third interval $(3 \le x \le 4)$, put $i=2$ in eq^n(i).
we get

$$f(x) = -\frac{6}{7}(4-x)^3 - \frac{6}{7}(x-3)^3 + \frac{13}{7}(4-x) - \frac{5}{7}(x-3)$$

For ~~third~~ fourth interval $(4 \le x \le 5)$, put $i=3$ in eq^n(i).
We get

$$f(x) = \frac{5}{7}(5-x)^3 - \frac{5}{7}(5-x) + (x-4)$$

The required cubic spline is:

$$\therefore \quad f(x) = \begin{cases} \frac{5}{7}(x-1)^3 + (2-x) - \frac{5}{7}(x-1), & 1 \le x \le 2 \\ \frac{5}{7}(3-x)^3 - \frac{6}{7}(x-2)^3 - \frac{5}{7}(3-x) + \frac{13}{7}(x-2), & 2 \le x \le 3 \\ -\frac{6}{7}(4-x)^3 - \frac{6}{7}(x-3)^3 + \frac{13}{7}(4-x) - \frac{5}{7}(x-3), & 3 \le x \le 4 \\ \frac{5}{7}(5-x)^3 - \frac{5}{7}(5-x) + (x-4), & 4 \le x \le 5. \end{cases}$$

Introduction of Regression:

Regression is a statistical Method that helps to estimate the relation and trend between two variables.

It can be used to predict the value of a dependent variable corresponding to a given value of an independent variable.

In regression process, we first plot the given data points in the graph and try to estimate a curve of best fit.

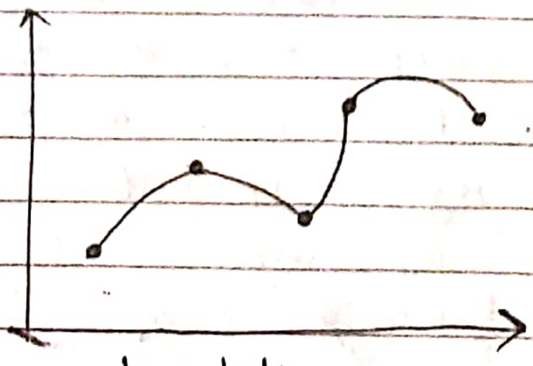
Conceptually the regression process is similar to interpolation, however it is a different approach. Interpolation is used to determine the value of dependent variable corresponding to the value of independent variable lying within a range ~~and we can use regression for the same~~. Interpolation is used when we have too few data points and the observations are not supposed to follow the similar trend.

Regression analysis is suitable to use if we have enough data points and the trend is supposed to be followed ahead. Interpolation is more accurate than regression but it will be too complex in case of large no. of data points. Regression is typically used for trend analysis.

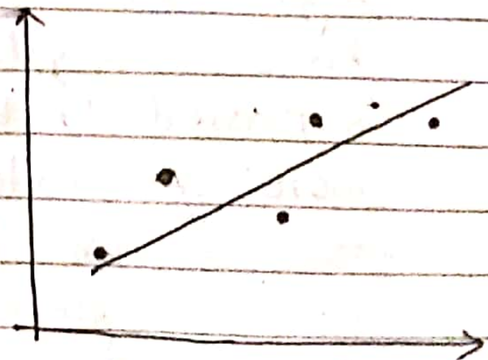
fig: Linear and non-linear regression

Interpolation Vs Regression

Interpolation	Regression
It is used when we have few observation points and we need to estimate the unknown value at another point	It is used when we have large observation points and we need to estimate the trend of variables.
It is more accurate in comparison to regression	It is less accurate in comparison to interpolation.
It is complex for large data points.	It is complex for less data points



Interpolation.



Regression for the same data set.

Least Square Method

The least square method is a form of regression analysis used to determine the line/curve of best fit for the given data points by minimizing the square of errors.

Mathematically, we try to minimize the sum of square of errors to get the curve of best fit.

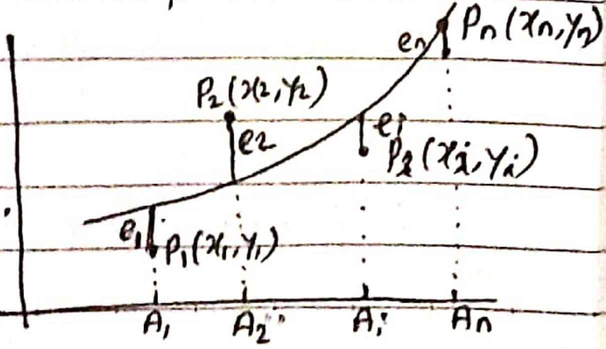
$$\sum e_i^2 = \sum (y_i - f(x_i))^2, \quad i = 1,2,\dots,n.$$

The errors in observation may be +ve or -ve. If we try to obtain the total error just by taking the sum of errors, the result may be zero, which is simply missleading result. According to the principle of least square, it is recomended to consider the sum of square of errors and try to minimize the result.

let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be n data points. We can fit the curve of the form,

$$y = f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n \quad \text{(i)}$$

The data points and the regression ~~points~~ curve can be shown as in the given graph.



let $y_1 = A_1B_1, y_2 = A_2B_2, \dots, y_n = A_nB_n$.

Let the curve of best fit be $y = f(x)$. At any point x_i, the observed value is y_i and the expected value is $f(x_i)$. The difference between observed and expected value is the error, given by

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

By using the least square principle, we try to find the curve of best fit such that the sum of e_i^2, $i = 1, 2, \ldots, n$ becomes minimum.

ie. $$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 \text{ is minimum.}$$

This technique of determining the curve of best fit by minimizing the sum of square of errors is known as principle of least square.

Linear Regression :-

Fitting a straight line is simplest approach of regression analysis. Linear regression is an approach of modeling the linear relationship between a dependent variable (y) with independent variable (x).

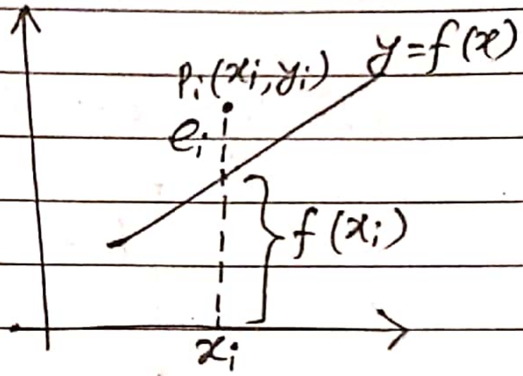
Linear regression are of two types based on the no. of independent variables.

a) Simple Linear Regression (Single independent variable)
b) Multiple Linear Regression (Many independent variables).

Simple Linear Regression

In simple linear regression, we try to estimate a straight line of best fit from the given observation points, having single independent variable and single dependent variable.

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the given observation points.



Let the line of best fit be.

$$f(x) = a + bx \quad \text{——— (i)}$$

The error in i^{th} observation is,

$$e_i = y_i - f(x_i)$$
$$= y_i - a - bx_i$$

The sum of square of errors is

$$E = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad \text{——— (ii)}$$

We must select/find regression coefficients a and b, such that E is minimum.
Necessary condition for E to be minimum is

$$\frac{\partial E}{\partial a} = 0 \quad \text{and} \quad \frac{\partial E}{\partial b} = 0$$

or, $$\frac{\partial E}{\partial a} = 2 \sum_{i=1}^{n} (y_i - a - bx_i) \cdot (-1) = 0$$

or, $$\sum_{i=1}^{n} y_i - na - b \sum_{i=1}^{n} x_i = 0 \quad \text{——— (iii)}$$

$$\therefore \boxed{a = \bar{y} - b\bar{x}}$$

classmate

Date

Page

Again $\frac{\partial E}{\partial b} = 0$

or, $2 \sum (y_i - a - bx_i) \cdot \frac{\partial}{\partial b}(y_i - a - bx_i) = 0$

or, $\sum (y_i - a - bx_i)(-x_i) = 0$

or, $-\sum x_i y_i + a\sum x_i + b \sum x_i^2 = 0$

or, $a \sum x_i + b \sum x_i^2 - \sum x_i y_i = 0$ ——— (iv)

Solving eqn (iii) & (iv), we get

$$\boxed{b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}}$$

Hence, we can use the following formulae to find the values of regression coefficients a & b.

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$a = \frac{\sum y_i}{n} - b \frac{\sum x_i}{n} = \bar{y} - b\bar{x}.$$

CSIT 066 Q. find the least square line that fits the following data.

x	1	2	3	4	5	6
$f(x)$	5.04	8.12	10.64	13.18	16.20	20.04

→ Here, $n = 6$

$\Sigma x =$ $\qquad$ $\Sigma x^2 =$

$\Sigma y =$

$\Sigma xy =$

use formula to find a & b. Ans " $y = a + bx$.

Q. find the least square regression line $y = a + bx$ from the following data & estimate y at $x = 10$.

x	0	1	2	3	4
y	2	3	5	4	6

Ans: $a = 2.2$

$b = 0.9$

$f(10) = 11.2$.

Non-Linear Regression

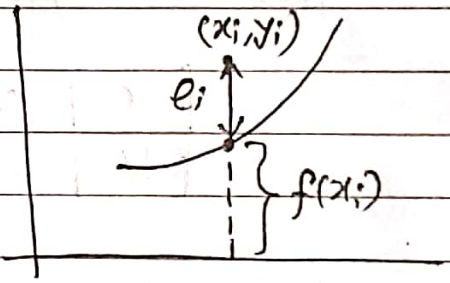
Non-linear regression is an approach of modeling the relationship between variables in terms of non-linear functions such as exponential & polynomial functions.

1. Non-linear Regression Using Exponential Model

Let us consider the n observation points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n). Let the observation pattern follow the exponential model of the form,

$$f(x) = a e^{bx} \quad \text{——— (i)}$$

The error in ith observation is,



$$e_i = y_i - f(x_i)$$
$$= y_i - a e^{bx_i}$$

The sum of square of errors is

$$E = \sum e_i^2$$

$$\therefore E = \sum (y_i - a e^{bx_i})^2 \quad \text{——— (ii)}$$

According to the least square principle, the exponential equation $y = ae^{bx}$ will be best fit curve if E is minimum. We have to find the values of regression coefficient a and b such that E is minimum. The necessary condition for E to be minimum is,

$$\frac{\partial E}{\partial a} = 0 \quad \text{and} \quad \frac{\partial E}{\partial b} = 0$$

Taking $\frac{\partial E}{\partial a} = 0$.

or, $\frac{\partial}{\partial a}\left(\sum (y_i - a e^{bx_i})^2\right) = 0$

or, $2\sum(y_i - a e^{bx_i})(-e^{bx_i}) = 0$

or, $$\boxed{a = \frac{\sum y_i e^{bx_i}}{\sum e^{2bx_i}}} \quad \text{—— (iii)}$$

Taking $\frac{\partial E}{\partial b} = 0$,

or, $\frac{\partial}{\partial b}\left(\sum (y_i - a e^{bx_i})^2\right) = 0$

or, $2\sum(y_i - a e^{bx_i})(-a x_i e^{bx_i}) = 0$

or, $$\boxed{\sum x_i y_i e^{bx_i} - a \sum x_i e^{2bx_i} = 0} \quad \text{—— (iv)}$$

Substitute the value of a from (iii) to (iv) and we can find b. With value of b, eqn (iii) gives a.

Note: Exponential regression model can be solved by converting the problem to linear regression. The curve (exponential) of best fit is;

$$f(x) = a e^{bx}$$

or, $y = a e^{bx}$

or $\log y = \log a + bx \quad (\log = \ln)$

Assume $z = \ln y$, $A = \ln a$ & $B = b$.

The eqn becomes $\boxed{z = A + Bx}$.

Now, use linear regression method to find A & B and hence a and b.

Q. Determine the equation of an exponential curve of best fit for the following data.

x	0	1	3	5	7	9
y	1	0.891	0.708	0.562	0.447	0.355

→ The exponential curve of best fit is given by

$$y = a e^{bx} \quad \text{—— (i)}$$

or, $\ln y = \ln a + bx$

Let $Y = \ln y$, $A = \ln a$ and $B = b$. Then the above equation reduces to

$$\therefore \quad Y = A + Bx$$

Here the equation $Y = A + Bx$ is a linear one. For the best fit, the linear regression coefficients A and B can be written as,

$$A = \frac{\sum Y_i}{n} - B\frac{\sum x_i}{n}$$

& $$B = \frac{n\sum x_i Y_i - \sum x_i \sum Y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Tabulating the values,

x	y	$Y = \ln y$	x^2	xY
0	1	0	0	0
1	0.891	-0.11541	1	-0.11541
3	0.708	-0.34531	9	-1.0359
5	0.562	-0.57625	25	-2.8813
7	0.447	-0.80520	49	-5.6364
9	0.355	-1.0356	81	-9.3207
$\sum x = 25$		$\sum Y = -2.8778$	$\sum x^2 = 165$	$\sum xY = -18.99$

$$B = \frac{6 \times (-18.99) - (25) \times (-2.8778)}{6 \times 165 - (25)^2}$$

$$= \frac{-113.94 + 71.945}{990 - 625}$$

$$= \frac{-41.995}{365}$$

or, $B = -0.1151$

$\therefore \boxed{b = -0.1151}$

Again,

$$A = \frac{\sum Y_i}{n} - B \frac{\sum x_i}{n}$$

$$= \frac{-2.8778}{6} - \frac{(-0.1151) \times 25}{6}$$

$$= -0.4796 + 0.4796$$

$$= 0$$

or, $\ln a = 0 \Rightarrow \boxed{a = 1}$

Hence the desired exponential curve of best fit is

$$y = a e^{bx}$$

$$\therefore \boxed{y = e^{-0.1151 x}}$$

Q. Fit the curve $y = ea^{bx}$ through the data given.

x	-4	-2	0	1	2	4
y	0.57	1.32	4.12	6.65	11	30.3

Ans: $y = 4.006\, e^{0.503 x}$.

2. Non-Linear Regression Using Polynomial Model

It is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled in the form of m^{th} degree polynomial.

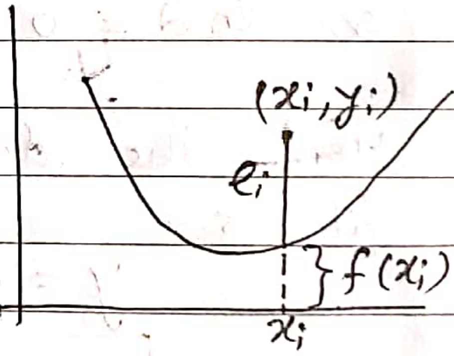
Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be the n data points given. We want to fit an m degree polynomial through the given data points.

The polynomial of degree m is of the form,

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m \quad \text{——(i)}$$

$$, m < n$$

The error in i^{th} observation is given by



$$e_i = y_i - f(x_i)$$
$$= y_i - a_0 - a_1 x_i - \dots - a_m x_i^m$$

Sum of square of the error is

$$E = \sum e_i^2$$
$$\therefore E = \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \dots - a_m x_i^m)^2$$

From least square principle, the eq^n (i) will be best fit if E is minimum. The regression coefficients $a_0, a_1, a_2, \dots, a_m$ can be obtained by solving $\frac{\partial E}{\partial a_0} = 0$, $\frac{\partial E}{\partial a_1} = 0$, $\frac{\partial E}{\partial a_2} = 0, \dots, \frac{\partial E}{\partial a_m} = 0$.

The solution for $a_0, a_1, a_2, \ldots, a_n$ can be obtained by solving the equations represented by the following matrix form.

$$\begin{bmatrix} n & \Sigma x & \Sigma x^2 & \cdots & \Sigma x^m \\ \Sigma x & \Sigma x^2 & \Sigma x^3 & \cdots & \Sigma x^{m+1} \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 & \cdots & \Sigma x^{m+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma x^m & \Sigma x^{m+1} & \Sigma x^{m+2} & \cdots & \Sigma x^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix} = \begin{bmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \\ \cdots \\ \cdots \\ \Sigma x^m y \end{bmatrix}$$

CSIT 069

Q. Find the best fitting quadratic polynomial from the following data using least square principle.

x	-2	-1.2	0	1	1.2	2.5	3	4.5	6.3
y	10.39	2.96	-2	-2.63	-2.46	0.83	3.1	12.8	30.4

→ The normal equations for the polynomial regression of degree 2 (quadratic polynomial) is given by:

$$\begin{bmatrix} n & \Sigma x & \Sigma x^2 \\ \Sigma x & \Sigma x^2 & \Sigma x^3 \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Sigma y \\ \Sigma xy \\ \Sigma x^2 y \end{bmatrix}$$

The quadratic polynomial of best fit is

$$y = a_0 + a_1 x + a_2 x^2 \quad \text{———} \quad (i)$$

Tabulating the given data,

x	y	x^2	x^3	x^4	xy	x^2y
-2	10.39	4	-8	16	-20.78	41.56
-1.2	2.96	1.44	-1.728	2.074	-3.552	4.2624
0	-2	0	0	0	0	0
1	-2.63	1	1	1	1	-2.63
1.2	-2.46	1.44	1.728	2.074	-2.952	-3.5424
2.5	0.83	6.25	15.625	39.0625	2.075	5.1875
3	3.1	9	27	81	9.3	27.9
4.5	12.8	20.25	91.125	410.063	57.6	259.2
6.3	30.4	39.69	250.047	1575.296	191.52	1206.576
Σx = 15.3	Σy = 53.39	Σx^2 = 83.07	Σx^3 = 376.797	Σx^4 = 2116.57	Σxy = 234.211	$\Sigma x^2 y$ = 1538.52

The normal equations becomes

$$\begin{bmatrix} 9 & 15.3 & 83.07 \\ 15.3 & 83.07 & 376.797 \\ 83.07 & 376.797 & 2116.57 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \cancel{15.3}\ 53.39 \\ 234.211 \\ 1538.52 \end{bmatrix}$$

Solving the above equations, we get

$a_0 = -1$

$a_1 = -2.45$

$a_2 = 1.20$

$\therefore$ The polynomial of best fit is

$$\boxed{y = -1 - 2.45x + 1.2x^2}$$

Q. Fit the quadratic curve through the following data set and estimate y at $x = 12$.

x	1	3	4	5	6	7	8	9	10
$f(x)$	2	7	8	10	11	11	10	9	8

→ Ans: $a_0 = -1.46$

$a_1 = 3.605$

$a_2 = -0.268$

polynomial: $\boxed{y = -1.46 + 3.605x - 0.268x^2}$