1. Introduction

Welcome to the series of E-learning modules on fitting of parabolic curves. In this section we find the method of fitting parabolic curves to the given data by the method of least squares and try to predict the future values.

By the end of this session, you will be able to:

a) Explain the method of fitting parabolic curve using the method of least squares
b) Predict the future values by assuming that the same trend will continue

While identifying the curve that best approximates the set of points, if we find the variation of parabolic type, we can fit a parabolic curve to the given data as explained below.

Let the equation to fit a parabolic curve y be given by the equation,
Y is equal to a plus b into x plus c into x square.

Principle of least squares consists of minimizing the sum of squared of the deviations between the given values of y and their estimates is given by the above equation.

In other words, we have to find a, b and c such that for given values of y corresponding to n different values of x.

That is E is equal to summation over i, y i minus b into x i minus c into x i square whole square is minimum.

For maxima or minima of E, for variations in a, b and c, we should have,
D E by d a is equal to zero is equal to minus 2 into summation y i minus a minus b into x i minus c into x i square
Implies, summation y i is equal to n into a plus b into summation x i plus c into summation x i square, which is the first normal equation.

D E by d b is equal to zero is equal to minus 2 into summation x i into y i minus a minus b into x i minus c into x i square
Implies, summation x i into y i is equal to a into summation x i plus b into summation x i square plus c into summation x i cube, the 2nd normal equation.
D E by d c is equal to zero is equal to minus 2 into summation x i square into y i minus a minus b into x i minus c into x i square
Implies, summation x i square into y i is equal to a into summation x i square plus b into summation x i cube plus c into summation x i power 4, the 3rd normal equation.

The values of summation y i, summation x i, summation x i square, summation x i cube, summation x i power 4, summation x i into y i, summation x i square into y i are obtained from the given data.
By substitution and simplification we can obtain a, b and c.
With these values of a, b and c, the parabolic curve is given by,
y is equal to a plus b into x plus c into x square.

2. Illustration 1

Illustration 1:
Fit a second degree parabola to the following data using the method of least squares. Estimate y when x is equal to 4 point zero.

Figure 1

X	Y
0.5	72
1.0	110
1.5	158
2.0	214
2.5	290
3.0	380

The parabola fitted to the above data is given by,
y is equal to a plus b into x plus c into x square.

And the normal equations are given by,
Summation y is equal to n into a plus b into summation x plus c into summation x square
Summation x into y is equal to a into summation x plus b into summation x square plus c into summation x cube and
Summation x square into y is equal to a into summation x square plus b into summation x cube plus c into summation x power 4.

To solve for a, b and c, we construct the following table.

Figure 2

X	Y	x=(X-1.75).4	x^2	x^3	x^4	xy	x^2y
0.5	72	-5	25	-125	625	-360	1800
1.0	110	-3	9	-27	81	-330	990
1.5	158	-1	1	-1	1	-158	158
2.0	214	1	1	1	1	214	214
2.5	290	3	9	27	81	870	2610
3.0	380	5	25	125	625	1900	9500
	1224	**0**	**70**	**0**	**1414**	**2136**	**15272**

As the given x values in decimals, for simplification, we do not use the given values as it is. We find the values of x such that, x is equal to x values minus 1 point 75 whole into 4.

Where 1 point 75 is the middle most values, so that summation x is equal to 0. The values of x are written in the 3rd column.

Forth column gives the x square values, which is obtained by squaring the numbers of column 3.

Fifth column is obtained by cubing the values of the column 3.

Sixth column is found by finding power 4 of values of the column 3.

Seventh column is obtained by multiplying the values of the column 3 and column 2.

And finally, the values of the column 8 are obtained by multiplying the values of the column 2 and 4.
Once we find the values of all the columns, finally we find the totals of all the columns, which are then substituted in the normal equations written above.

The first normal equation is,
1224 is equal to 6 into a plus 0 into b plus 70 into c
Implies 1224 is equal to 6 into a plus 70 into c

The second normal equation is,
2136 is equal to zero into a plus 70 into b plus zero into c
That is, 2136 is equal to 70 into b
Implies, b is equal to 2136 divided by 70 is equal to 30 point 51.

The third normal equation is,
15272 is equal to 70 into a plus 0 into b plus 1414 into c
Implies, 15272 is equal to 70 into a plus 1414 into c.

To find a and c we solve the equations 1 and 3 as follows.

Consider the determinant delta is equal to 6, 70, 70, 1414 is equal to 6 into 1414 minus 70 into 70 is equal to 3584.

Delta 1 is equal to 1224, 70, 15272, 1414 is equal to 1224 into 1414, minus 15272 into 70 is equal to 6 lakh 61 thousand 696.

Delta 2 is equal to 6, 1224, 70, 15272 is equal to 6 into 15272 minus 70 into 1224 is equal to 5952.

Therefore, a is equal to delta 1 by delta is equal to 6 lakh 61 thousand 696 divided by 3584 is equal to 184 point 625 and c is equal to delta 2 divided by delta is equal to 5952 divided by 3584 is equal to 1 point 66.

Therefore, the parabolic equation fitted to the given data is given by,
Y is equal to 184 point 625 plus 30 point 51 into x plus 1 point 66 x square.

The estimated value of y when x is equal to 4 point 0, that is x is equal to 4 point 0 minus 1 point 75 into 4 is equal to 9 is given by,

Y is equal to 184 point 625 plus 30 point 51 into 9 plus 1 point 66 into 9 square
Is equal to 593 point 7.

3. Illustration 2

Compute the estimated values by fitting the second degree parabola to the given data in the below table. Also estimate the import for the year 2013.

Figure 3

Year	Imports ('000 tons)	Year	Imports ('000 tons)
2001	98	2006	45
2002	87	2007	57
2003	62	2008	96
2004	47	2009	100
2005	54		

The parabola fitted to the above data is given by,
y is equal to a plus b into x plus c into x square.

And the normal equations are given by,
Summation y is equal to n into a plus b into summation x plus c into summation x square

Summation x into y is equal to a into summation x plus b into summation x square plus c into summation x cube and

Summation x square into y is equal to a into summation x square plus b into summation x cube plus c into summation x power 4.

To solve for a, b and c, we construct the following table.

Figure 4

Year	Imports ('000 tons)	x= Year - 2005	x^2	x^3	x^4	xy	x^2y
2001	98	-4	16	-64	256	-392	1568
2002	87	-3	9	-27	81	-261	783
2003	62	-2	4	-8	16	-124	248
2004	47	-1	1	-1	1	-47	47
2005	54	0	0	0	0	0	0
2006	45	1	1	1	1	45	45
2007	57	2	4	8	16	114	228
2008	96	3	9	27	81	288	864
2009	100	4	16	64	256	400	1600
Total	**646**	**0**	**60**	**0**	**708**	**23**	**5383**

Since we have odd number of observations, the middle most value is 2005. Hence, to get the values of x we find the deviations of years from 2005.
The third column in the table gives these values.

The fourth column is computed by squaring the values of the third column that is minus 4 square is equal to 16, minus 3 square is equal to 9. Similarly all the other values are squared.

The fifth column is found by cubing the values of the 3rd column. That is minus 4 cube is equal to minus 64, minus 3 cube is equal to minus 27. Similarly all the other values are calculated.

The sixth column is computed by finding the forth power of the elements in the 3rd column. That is minus 4 power 4 is equal to 256, minus 3 power 4 is equal to 81. Similarly all the other values are calculated.

Seventh column is found by multiplying the elements in 2nd and third column. That is 98 into minus 4 is equal to minus 392, 87 into minus 3 is equal to minus 261 etc.

And the last column is obtained by multiplying the elements in 2nd and 4th column. That is 98 into 16 is equal to 1568, 87 into 9 is equal to 783 etc.
Once we find all the values, find the total of all the columns.

The first normal equation is given by,
646 is equal to 9 into a plus zero into b plus 60 into c
Implies, 646 is equal to 9 into a plus 60 into c

The second normal equation is given by,
23 is equal to 0 into a plus 60 into b plus 0 into c
That is 23 is equal to 60 into b
Implies, b is equal to 23 by 60 is equal to zero point 38.

The third normal equation is given by,
5383 is equal to 60 into a plus zero into b plus 1468 into c
Implies, 5383 is equal to 60 into a plus 708 into c

3. Illustration 2 (Contd.)

To find a and c we solve the equations using Cramer's rule as follows.

Consider the following determinants.
Delta is equal to determinant of 9, 60, 60, 708 is equal to 9 into 708 minus 60 into 60 is equal to 2772.

Delta 1 is equal to 646, 60, 5383, 708 is equal to 646 into 708 minus 5383 into 60 is equal t 1 lakh 34 thousand 388.

Delta 2 is equal to 9, 646, 60, 5383 is equal to 9 into 5383 minus 60 into 646 is equal to 9687.

Therefore, a is equal to delta 1 divided by delta is equal to 1 lakh 34 thousand 388 divided by 2 thousand 772 is equal to 48 point 48.

And c is equal to delta 2 divided by delta is equal to 9 thousand 687 divided by 2 thousand 772 is equal to 3 point 49.

Therefore, 48 point 48 plus zero point 38 into x plus 3 point 49 into x square.

To find the estimated imports in thousand tons corresponding to different years we substitute the corresponding values of x as indicated in the third column.

For the year 2001, x is equal to minus 4. Hence estimated import is y is equal to 48 point 48 plus 0 point 38 into minus 4 plus 3 point 49 into minus 4 square is equal to 102 point 8.

For the year 2002, x is equal to minus 3. Hence, y is equal to 48 point 48 plus 0 point 38 into minus 3 plus 3 point 49 into minus 3 square is equal to 78 point 77.

For the year 2003, x is equal to minus 2. Hence, y is equal to 48 point 48 plus 0 point 38 into minus 2 plus 3 point 49 into minus 2 square is equal to 61 point 68.

Proceeding like this we can obtain estimated values for all the years. That is estimated value for all the years.

That is for 2004, y is equal to 51 point 59
for 2005, y is equal to 48 point 48
for 2006, y is equal to 52 point 35
for 2007, y is equal to 63 point 20
for 2008, y is equal to 81 point 03
For 2009, y is equal to 105 point 84
To estimate the import for the year 2013, we take x is equal to 2013 minus 2005 is equal to 8.

Therefore, y is equal to 48 point 48 plus 0 point 38 into 8 plus 3 point 49 into 8 square is equal to, 274 point 88 thousand tons.

5. Illustration 3

Following table gives the population of India in crores over the years. Fit a parabolic curve to the data and estimate the population for the year 2 thousand 21.

Figure 5

Year	Population (Crores)	Year	Population (Crores)
1901	238.4	1961	439.23
1911	252.09	1971	84.16
1921	251.32	1981	683.33
1931	278.98	1991	864.42
1941	318.16	2001	1028074
1951	316.09	2011	1210.19

The parabola fitted to the above data is given by,
y is equal to a plus b into x plus c into x square.

And the normal equations are given by,
Summation y is equal to n into a plus b into summation x plus c into summation x square.

Summation x into y is equal to a into summation x plus b into summation x square plus c into summation x cube and

Summation x square into y is equal to a into summation x square plus b into summation x cube plus c into summation x power 4.

To solve for a, b and c, we construct the following table.

Figure 6

Year	Population (Crores)	x= (Year – 1956)/5	x^2	x^3	x^4	xy	x^2y
1901	2.38	-11	121	-1331	14641	-26.18	287.98
1911	2.52	-9	81	-729	6561	-22.68	204.12
1921	2.51	-7	49	-343	2401	-17.57	122.99
1931	2.79	-5	25	-125	625	-13.95	69.75
1941	3.18	-3	9	-27	81	-9.54	28.62
1951	3.16	-1	1	-1	1	-3.16	3.16
1961	4.39	1	1	1	1	4.39	4.39
1971	5.84	3	9	27	81	17.52	52.56
1981	6.83	5	25	125	625	34.15	170.75
1991	8.64	7	49	343	2401	60.48	423.36
2001	10.29	9	81	729	6561	92.61	833.49
2011	12.1	11	121	1331	14641	133.1	1464.1
Total	**64.63**	**0**	**572**	**0**	**48620**	**249.17**	**3665.27**

Since the values taken by x, that is years are very large, we consider the values of x by taking the deviation such that summation x becomes zero.
That is x is equal to year minus 1956 whole divided by 5.
That is x is equal to 1901 minus 1956 is equal to minus 11,
1911 minus 1956 is equal to minus 9 etc. These values are written in the column number 3.

Fourth column gives the square of values of the third column. That is minus 11 square is equal to 121, minus nine square is 81. Similarly all the other values are squared.

Fifth column is obtained by cubing the values of the third column. That is minus 11 cube is equal minus 1331, minus 9 cube is equal to minus 729. Similarly all the other values are calculated.

Sixth column is obtained by finding power 4 of the values of the third column. That is, minus 11 power 4 is equal to 14641, minus 9 power 4 is equal to 6561 etc.

Seventh column is given by multiplying the values of the column 3 and 2. That is minus 11 into 2 point 38 is equal to minus 26 point 18, minus 9 into 2 point 52 is equal to minus 22 point 68 etc.

The last column is obtained by multiplying the elements of 4^{th} and 2^{nd} columns. That is 121 into 2 point 38 is equal to 287 point 98, 81 into 2 point 52 is equal to 204 point 12 etc.

Once we find all the elements in different columns, we find the totals of all the columns, which are indicated by bold numbers in the table.

The first normal equal is 64 point 63 is equal to 12 into a plus 572 into c.

Second normal equation is given by, 249 point 17 is equal 572 into b, implies, b is equal to 249 point 1 seven divided by 572 is equal to zero point 4 three.

Third normal equation is given by, 3665 point 27 is equal to 572 into a plus 48 thousand 620 into c.

To find a and c we solve the equations 1 and 3 as follows.

A is equal to delta 1 by delta is equal to determinant of 64 point 63, 572, 3665 point 27, 48 thousand 620 divided by determinant of 12, 572, 572, 48620.
On simplification we get, 4 point 08.

Similarly, c is equal to delta 2 divided by delta is equal to 12, 64 point 63, 572, 3665 point 27 divided by determinant of 12, 572, 572, 48 thousand 620.
On simplification, we get, 0 point 027.

Therefore, the parabolic equation is fitted to the given data is given by,
y is equal to 4 point 08 plus zero point 43 into x plus 0 point 027 into x square.

Assuming that the same trend continues for another 10 years, the population of India for the year 2021 is obtained by putting x is equal to 2021 minus 1956 divided by 5, is equal to 13.

Therefore, the estimated population for the year 2021 is given by,
Y is equal to 4 point 08 plus 0 point 43 into 13 plus 0 point 027 into 13 square is equal to 14 point 233 crores.

Here's a summary of our learning in this session,

- The method of fitting parabolic curve using the method of least squares.
- Predict the future values by assuming that the same trend will continue.