LINEAR REGRESSION ANALYSIS

MODULE – XII

Lecture - 35

Polynomial Regression Models

Dr. Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

A model is said to be linear when it is linear in parameters. So the model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

and

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

are also the linear model. In fact, they are the second order polynomials in one and two variables respectively.

The polynomial models can be used in those situations where the relationship between study and explanatory variables is curvilinear. Sometimes a nonlinear relationship in a small range of explanatory variable can also be modeled by polynomials.

Polynomial models in one variable

The k^{th} order polynomial model in one variable is given by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon.$$

If $x_j = x^j,\ j = 1, 2, \ldots, k,$ then the model is multiple linear regressions model in *k* explanatory variables $x_1, x_2, \ldots, x_k$.

So the linear regression model $y = X\beta + \varepsilon$ includes the polynomial regression model. Thus the techniques for fitting linear regression model can be used for fitting the polynomial regression model.

For example:

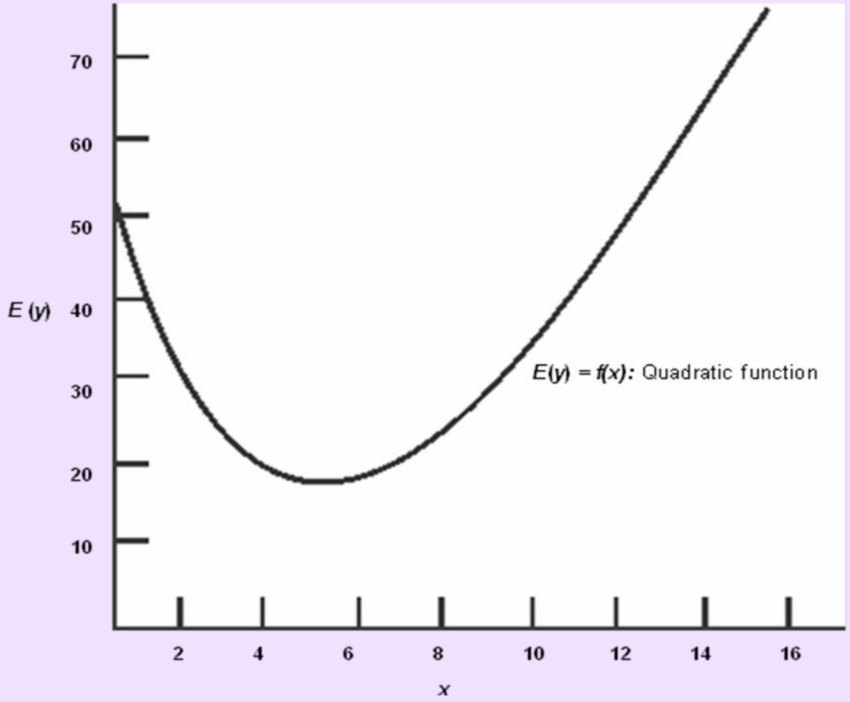
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

or

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

is a polynomial regression model in one variable and is called as **second order model** or **quadratic model.** The coefficients β_1 and β_2 are called the **linear effect parameter** and **quadratic effect parameter** respectively.

The interpretation of parameter β_0 is $\beta_0 = E(y)$ when $x = 0$ and it can be included in the model provided the range of data includes $x = 0$. If $x = 0$ is not included, then β_0 has no interpretation. An example of quadratic model is like as follows:



The polynomial models can be used to approximate a complex nonlinear relationship. The polynomial models is just the Taylor series expansion of the unknown nonlinear function in such a case.

Considerations in fitting polynomial in one variable

Some of the considerations in fitting polynomial model are as follows:

1. Order of the model

The order of the polynomial model is kept as low as possible. Some transformations can be used to keep the model to be of first order. If this is not satisfactory, then second order polynomial is tried. Arbitrary fitting of higher order polynomials can be a serious abuse of regression analysis. A model which is consistent with the knowledge of data and its environment should be taken into account. It is always possible for a polynomial of order $(n-1)$ to pass through n points so that a polynomial of sufficiently high degree can always be found that provides a "good" fit to the data. Such models neither enhance the understanding of the unknown function nor be a good predictor.

2. Model building strategy

A good strategy should be used to choose the order of an approximate polynomial.

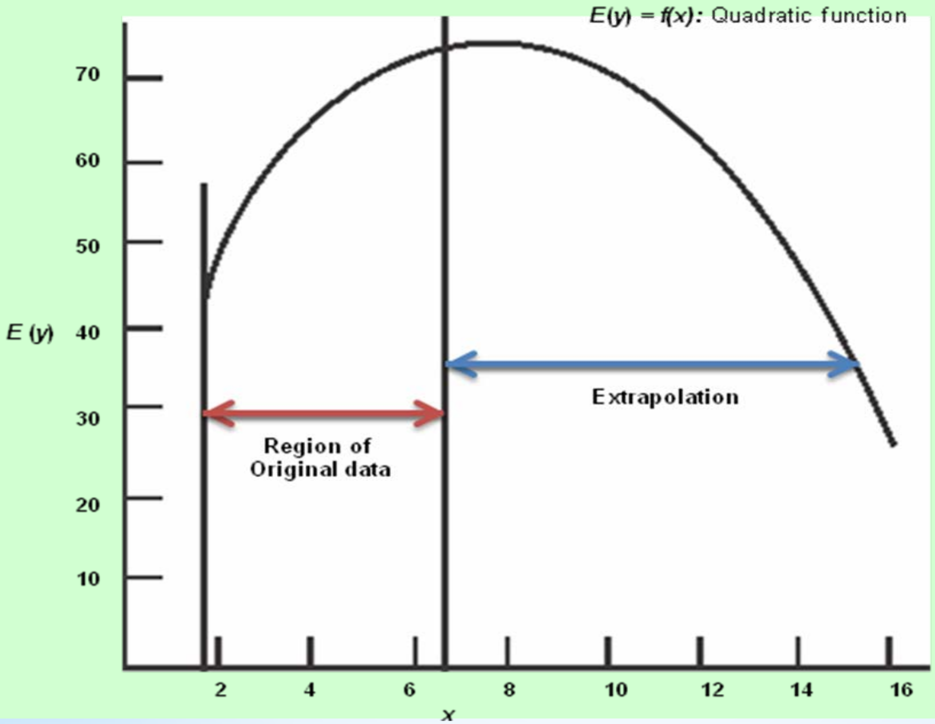
One possible approach is to successively fit the models in increasing order and test the significance of regression coefficients at each step of model fitting. Keep the order increasing until t-test for the highest order term is nonsignificant. This is called as **forward selection procedure.**

Another approach is to fit the appropriate highest order model and then delete terms one at a time starting with highest order. This is continued until the highest order remaining term has a significant t-statistic. This is called as **backward elimination** procedure.

The forward selection and backward elimination procedures does not necessarily lead to same model. The first and second order polynomials are mostly used in practice.

3. Extrapolation

One has to be very cautious in extrapolation with polynomial models. The curvatures in the region of data and region of extrapolation can be different. For example, in the following figure, the trend of data in the region of original data is increasing but it is decreasing in the region of extrapolation. So predicted response will not be based on the true behaviour of the data.



In general, polynomial models may have unanticipated turns in inappropriate directions. This may provide incorrect inferences in interpolation as well as extrapolation.

4. Ill-conditioning

A basic assumption in linear regression analysis is that X-matrix is of full column rank. In polynomial regression models, as the order increases, the $X'X$ matrix becomes ill-conditioned. As a result, the matrix $(X'X)^{-1}$ may not be accurate and the parameters will be estimated with considerable error.

If values of x lie in a narrow range then the degree of ill-conditioning increases and multicollinearity in the columns of X matrix enters. For example, if x varies between 2 and 3 , then x^2 varies between 4 and 9. This introduces strong multicollinearity between x and x^2.

5. Hierarchy

A model is said to be hierarchical if it contains the terms x, x^2, x^3, etc. in a hierarchy. For example, the model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \varepsilon$$

is hierarchical as it contains all the terms upto order four. The model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_4 x^4 + \varepsilon$$

is not hierarchical as it does not contain the term of x^3.

It is expected that all polynomial models should have this property because only hierarchical models are invariant under linear transformation. This requirement is more attractive from mathematics point of view. In many situations, the need of model may be different. For example, the model

$$y = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2 + \varepsilon$$

needs a two factor interaction which is provided by the cross-product term. A hierarchical model would need inclusion of x_2 which is not needed from the point of view of statistical significance perspective.

Orthogonal polynomials

While fitting a linear regression model to a given set of data, we begin with simple linear regression model. Suppose later we decide to change it to a quadratic or wish to increase the order from quadratic to a cubic model etc. In each case, we have to begin the modeling from scratch, i.e., from simple linear regression model. It would be preferable to have a situation in which adding an extra term merely refine the model in the sense that by increasing the order, we do not need to do all the calculations from the scratch. This aspect was of more importance in pre-computer era when all the calculations were done manually. This cannot be achieved by using the powers $x^0 = 1, x, x^2, x^3 \ldots$ in succession. But it can be achieved by a system of orthogonal polynomials. The k^{th} orthogonal polynomial has degree k. Such polynomials may be constructed by using Gram-Schmidt orthogonalization .

Another issue in fitting the polynomials in one variable is ill conditioning. An assumption in usual multiple linear regression analysis is that all the independent variables are independent. In polynomial regression model, this assumption is not satisfied. Even if the ill-conditioning is removed by centering, there may still exist high levels of multicollinearity. Such difficulty is overcome by orthogonal polynomials.

The classical cases of orthogonal polynomials of special kinds are due to Legendre, Hermite and Tehebycheff polynomials. These are **continuous orthogonal polynomials** (where the orthogonality relation involve integrating) whereas in our case, we have **discrete orthogonal polynomials** (where the orthogonality relation involves summation).

Analysis

Consider the polynomial model of order *k* is one variable as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_k x_i^k + \varepsilon_i, \quad i = 1, 2, ..., n.$$

When writing this model as

$$y = X\beta + \varepsilon,$$

the columns of X will not be orthogonal. If we add another terms $\beta_{k+1} x_i^{k+1}$, then the matrix $(X'X)^{-1}$ has to be recomputed and consequently, the lower order parameters $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ will also change.

Consider the fitting of following model:

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \alpha_2 P_2(x_i) + ... + \alpha_k P_k(x_i) + \varepsilon_i, \; i = 1, 2, ..., n$$

where $P_u(x_i)$ is the *u*th order orthogonal polynomial defined as

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \; r \neq s, r, s = 0, 1, 2, ..., k$$

$$P_0(x_i) = 1.$$

In the context of $y = X\beta + \varepsilon,$ the *X*-matrix in this case is given by

$$X = \begin{bmatrix} P_0(x_1) & P_1(x_1) & \cdots & P_k(x_1) \\ P_0(x_2) & P_1(x_2) & \cdots & P_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(x_n) & P_1(x_n) & \cdots & P_k(x_n) \end{bmatrix}.$$

Since this *X*-matrix has orthogonal columns, so *X'X* matrix becomes

$$X'X=\begin{bmatrix} \sum_{i=1}^{n} P_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} P_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} P_k^2(x_i) \end{bmatrix}.$$

The ordinary least squares estimator is $\hat{\alpha}=(X'X)^{-1}X'y$ which for α_j is

$$\hat{\alpha}_j=\frac{\sum_{i=1}^{n} P_j(x_i)y_i}{\sum_{i=1}^{n} P_j^2(x_i)},\quad j=0,1,2,\ldots,k$$

and its variance is obtained from $V(\hat{\alpha})=\sigma^2(X'X)^{-1}$ as

$$Var(\hat{\alpha}_j)=\frac{\sigma^2}{\sum_{i=1}^{n}\left[P_j(x_i)\right]^2}.$$

When σ^2 is unknown, it can be estimated from analysis of variance table.

Since $P_0(x_i)$ is a polynomial of order zero, set it as $P_0(x_i)=1$ and consequently

$$\hat{\alpha}_0=\hat{y}=\overline{y}.$$

The residual sum of squares is

$$SS_{res}(k) = SS_T - \sum_{j=1}^{k} \hat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i) y_i \right].$$

The regression sum of squares is

$$SS_{reg}(\hat{\alpha}_j) = \hat{\alpha}_j \sum_{i=1}^{n} P_j(x_i) y_i$$

$$= \frac{\left[\sum_{i=1}^{n} P_j(x_i) y_i \right]^2}{\sum_{i=1}^{n} P_j^2(x_i)}.$$

This regression sum of squares does not depend on other parameters in the model.

The analysis of variance table in this case is given as follows

Source of variation	Degrees of freedom	Sum of squares	Mean squares
$\hat{\alpha}_0$	1	$SS(\hat{\alpha}_0)$	-
$\hat{\alpha}_1$	1	$SS(\hat{\alpha}_1)$	$SS(\hat{\alpha}_1)$
$\hat{\alpha}_2$	1	$SS(\hat{\alpha}_2)$	$SS(\hat{\alpha}_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\hat{\alpha}_k$	1	$SS(\hat{\alpha}_k)$	$SS(\hat{\alpha}_k)$
Residual	$n-k-1$	$SS_{res}(k)$ (by subtraction)	SS_{res}
Total	n	SS_T	

If we add another term $P_{k+1}(x_i)\alpha_{k+1}$ in the model, then the model is

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + ... + \alpha_{k+1} P_{k+1}(x_i) + \varepsilon_i;\ i = 1, 2, ..., n$$

and then we just need $\hat{\alpha}_{k+1}$ which can be obtained as

$$\hat{\alpha}_{k+1} = \frac{\sum_{i=1}^{n} P_{k+1}(x_i) y_i}{\sum_{i=1}^{n} \left[P_{k+1}(x_i) \right]^2}.$$

Notice that:

- We need not to bother for other terms in the model.
- Simply concentrate on the newly added term only.
- No re-computation of $(X'X)^{-1}$ or any other $\hat{\alpha}_j (j \neq k+1)$ is necessary due to orthogonality of polynomials.
- Thus higher order polynomials can be fitted with ease.
- Terminate the process when a suitably fitted model is obtained.