3/3/20

UNIT - 4

Polynomial regression models:

The linear regression model $y = X\beta + \varepsilon$, is a general model for fitting any relationship that is linear in the unknown parameter β. This includes the important class of polynomial regression models..

For Ex: The 2^{nd}-order polynomial in one independent variable.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \quad \text{——— (1)}$$

and the second-order polynomial in two indept. variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad \text{——— (2)}$$

are regression models.

Polynomials are widely used in situations where the response is curvilinear, as even complex non-linear relationships can be adequately modeled by polynomials over reasonably small range of the x's.

Polynomial models in one variable:

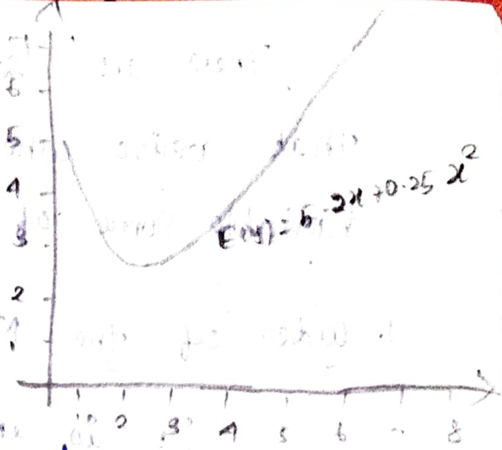
Basic Principles:

As an example of a polynomial regression model in one variable, consider

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

This model is called a second-order model in one variable. It is also sometimes called a quadratic model, since the expected value of y is,

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



Which describes a quadratic function. A typical example is shown in figure. we often call β_1 the linear effect parameter and β_2 the quadratic effect parameter.

The parameter β_0 is the mean of y when $x=0$ if the range of the data includes $x=0$, otherwise β_0 has no physical interpretation.

In general, the k^{th} order polynomial model in one variable in one variable is

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon \quad \text{③}$$

If we set $x^j = x_j$, $j = 1, 2, \dots k$. then equ ③ becomes a multiple linear regression model in the k regressions $x_1, x_2, \dots x_k$. Thus, a polynomial model of order k may be fitted using the techniques studied previously.

Polynomial models are useful in situations where the analyst knows that curvilinear effects are present in the true response function. They are also useful as approximating functions to unknown and possibly very complex non-linear relationship.

In this sense, the polynomial model is just the Taylor series expantation of the unknown function. This type of application seems to occur most often in practice.

There are several important considerations that arise when fitting a polynomial in one variable. Some of these are discussed below.

1. Order of the Model:

It is important to keep the order of the model as low as possible when the response function appears to be curvilinear, transformations should be tried to keep the model 1st-order. of this fails a 2nd-order polynomial should be tried. As a general rule the use of high-order polynomial ($k>2$) should be avoided unless they can be justified for reasons outside the data. A low-order model in a transformed variable is almost always preferable to a high-order model in the original matrix. Arbitrary fitting of high order polynomial is a serious abuse of regression analysis. One should always maintain a sense of parsimony, that is use, the simplest possible model that is consistent with the data and knowledge of the problem environment. Remember that in an extreme case it is always possible to pass a ploynomial, order n-1 through 'n' points so that a polynomial of sufficiently high degree can always be found that provides a good fit to the data. Such a model would do nothing to enhance understanding of the unknown function, nor will it likely be a good predictor.

2. Model Building Strategy:

Various Strategies for choosing the order of an approximating Polynomial have been Suggested One apporach is to Successively fit models of increasing order until the 't' test for the highest order term is non-significantly. An alternate Procedure is to fit the highest order model appropriate and them delete terms one at a time. Starting with the highest order until, the highest order remaining term has a significant 't' Statistics. These two procedures are called formal Selection and backword elimination, respectively. They do not neccessarily lead to the same model. In light of the comment in 1 above, these procedures Should be used carefully. In most situations we Should restrict our attention to 1st and 2nd order polynomials.

3. Extrapolation:

Extrapolation with polynomial models can be extremely hazardous. For example, consider the 2nd order model is if we extrapolate beyond the range of the original data. The predicted response fiams downward. This may be at adds with the true behaviour of the system. In general polynomial models may term in unanticipated and enapproapriate directions, both in enterpolation and in extrapolation.

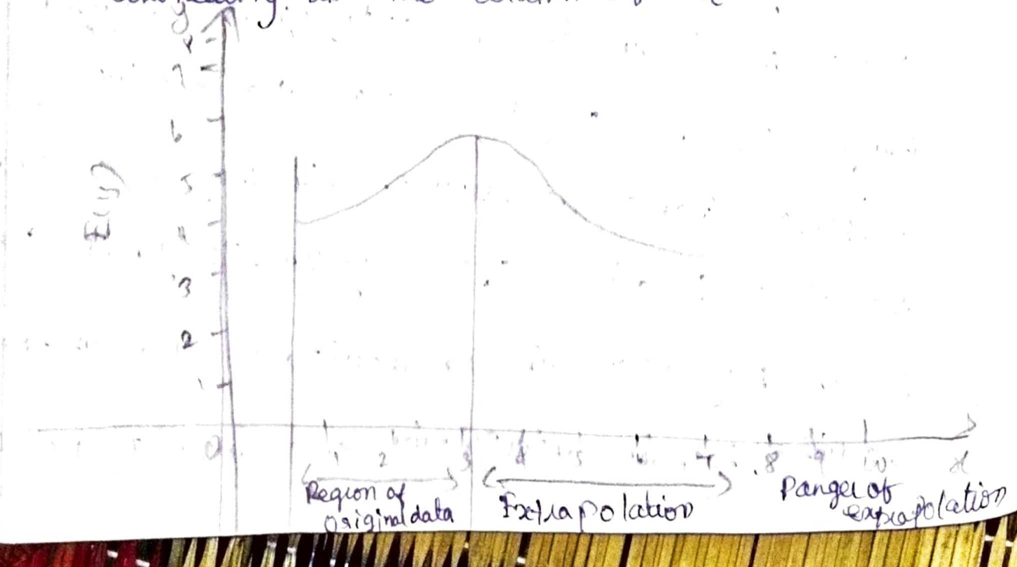
4. Ill - condition ? :

As the order of the polynomial increases the $x'x$ matrix becomes ill conditioned. This means that the matrix inversion calculation will be in correct and considerable error may be introduced into two parameter estimator.

For ex: see forsythe (1957) non - essential ill - condition caused by the arbitrary choice of origin can be removed by first centering the regressor variables (i.e. correcting x for its average $\bar{x}$) but as Bradly and Srivastava [1979) point out even centering the data can still result, in large sample correlations between certain regression co-efficients. A method for dealing with this problem will be discussed in section 7.4.

5. Ill - condition II :

If the values of x are limited to a narrow range, there can be significant ill - conditioning or multicollinearity in the column of the x matrix.



For example, if x varies between 1 and 2, x^2 varies b/w 1 and 4, which could create strong multicollinearity b/w x and x^2.

6. Hierarchy:

The regression model, $y = \beta_0 + \beta_1 x + \beta_2 x_2^2 + \beta_3 x^3 + \epsilon$. is said to be hierarchical because it contains all terms of three and lower. By contrast, the model

$$y = \beta_0 + \beta_1 x + \beta_3 x^3 + \epsilon.$$

is not hierarchical. Peixoto [1987, 1990] points out that only hierarchical models are invariant under linear transformations and suggests that all polynomial models should have this property (the phrase "a hierarchically well-formulated model" is frequently used). We have mixed feelings about this as a hard-and-fast rule. It is certainly attractive to have the model from preserved following a linear transformation (such as fitting the model in coded variables and then converting to a model in the natural variables), but it is purely a mathematical nicety. There are many mechanistic models that are not hierarchical.

For example; Newton's law of gravity is an inverse square law and the magnetic dipole law is an inverse cube law. Furthermore there are many situations in using a polynomial regression model to represent the results of a designed experiment where a model Such as.

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

would be supported by the data, where the cross product term represents a two factor interaction. Now, a hierarchical model would require the inclusion of the other main effects x_2. However, this other term could really be entirely unnecessary from a Statistical significance perspective. It may be perfectly logical from the viewpoint of the underlying Science or engineering to have an interaction in the model without one (or even in some cases either) of the individual main effects. This occurs frequently when some of the variables involved in the interaction are categorical. The best advice is to fit a model that has all terms significant, and to use discipline knowledge rather than an arbitary rule as an additional guide in model formulation.

Piecewise Polynomial Fitting (Splines).

Sometimes we find that a low order Polynomial provides a poor fit to the data and increasing the order of the polynomial modestly does not substantially improve the situation. Symptoms of this are the failure of the residual Sum of Square to Stabilize or residual plot that exhibit remaining unexplained Structure. This Problem may occur when the function behaves differently in different parts of the range of x occasionally transformations on x and/or y will eliminate this Problem. The usual approach, however, is to divide the range of x into segments and fit an appropriate curve in each segment. Spline functions offer a useful way to Perform this type of piece wise Polynomial fitting.

Splines are piecewise polynomials of order k. The joint points of the pieces are usually called knots. Generally, we require the function values and the 1^{st} $k-1$ derivatives to agree at the knots, so that the Spline is a continuous function with $k-1$ continuous derivatives. The cubic spline $(k=3)$ is usually adequate for most Practical Problems.

A cubic Spline with h knots, $t_1 < t_2 < \dots < t_h$ with continuous 1^{st} and 2^{nd} derivatives. can be written as,

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{h} \beta_i (x - t_i)_+^3$$

where,

$$(x - t_i)_+ = \begin{cases} (x_i - t_i) & ; \text{ if } x - t_i > 0 \\ 0 & ; \text{ if } x - t_i \leq 0 \end{cases} \quad \text{——} (4)$$

We assume that the positions of the knots are known. If the knot positions are parameters to be estimated, the resulting Problem is a non-linear regression problem. When knots

Positions are known, however, fitting equ (A) can be accomplished by a straight forward application of linear least squares.

Deciding on the number and position of the knots and the order of the polynomial in each segment is not simple. Wold [1974] suggests that there should be as few knots as possible, with at least four or five data points per segment. Considerable caution should be exercised here because the great flexibility of spline functions makes it very easy to "overfit" the data. (maximum or minimum) and one point of inflection per segment. Insofar as possible, the extreme points should be centered in the segment and the points of inflexion should be near the knots. When prior information about the data-generating process is available, this can sometimes aid in knot positioning.

The basic cubic spline mode (A) can be easily modified to fit polynomials of different order in each segment and to impose different continuity restrictions at the knots. If all h+1 polynomial pieces are of order 3, then a cubic spline model with no continuity restrictions is

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{h} \sum_{j=0}^{3} \beta_{ij} (x - t_i)_+^j$$

Where, $(x-t)^0_+$ equals 1 if $x \geq t$ and 0 if $x < t$. Thus, if a term $\beta_{ij}(x-t_i)^j_+$ is in the model, this forces a discontinuity at t_i in the j^{th} derivative of $S(x)$. If this term is absent, the j^{th} derivative of $S(x)$ is continuous at t_i. The fewer continuity restrictions required, the worse is the fit but the smoother the final curve will be. Determining both the order of the polynomial segments and the continuity restrictions that do not substantially degrade the fit can be done using standard multiple regression hypothesis-testing methods.

As an illustration consider a cubic spline with a single knot at t and no continuity restrictions; for example;

$$E(y) = S(x) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 +$$
$$\beta_{10}(x-t)^0_+ + \beta_{11}(x-t)^1_+ + \beta_{12}(x-t)^2_+ + \beta_{13}(x-t)^3_+$$

Note, that neither $S(x)$, $S'(x)$, nor $S''(x)$ is necessarily continuous at t because of the presence of the terms involving β_{10}, β_{11} and β_{12} in the model. To determine whether imposing continuity restrictions reduces the quality of the fit, test the hypotheses $H_0: \beta_{10} = 0$ [continuity of $S(x)$ $S'(x)$]. $H_0: \beta_{10} - \beta_{11} = 0$ [continuity of $S(x)$, and $S'(x)$]

$H_0: \beta_{10} = \beta_{11} = \beta_{12} = 0$ [continuity of $S(x)$, $S'(x)$, $S''(x)$]. To determine whether the cubic spline fits the data better than a single cubic polynomial over the range of x, simply test

$H_0: \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = 0$.

An excellent description of this approach to fitting splines is in Smith [1979]. A potential disadvantage of this model is that the $X'X$ matrix becomes ill-conditioned if there are a large no. of knots. This problem can be overcome by using different representation of the spline called the cubic B-spline. The cubic B-splines are defined in terms of divided differences,

$$B_i(x) = \sum_{j=i-4}^{i} \left[\frac{(x-t_j)_+^3}{\prod_{\substack{m=i-4 \\ m \neq j}}^{i} (t_j - t_m)} \right]; \quad i = 1, 2, \ldots, h+4 \qquad \text{(B)}$$

and,

$$E(y) = S(x) = \sum_{i=1}^{h+4} \gamma_i B_i(x) \qquad \text{(C)}$$

where, γ_i, $i = 1, 2, \ldots h+4$, are parameters to be estimated. In equ (B) there are eight additional knots, $t_{-3} < t_{-2} < t_{-1} < t_0$ and $t_{h+1} < t_{h+2} < t_{h+3} < t_{h+4}$. We usually take $t_0 = x_{min}$ and $t_{h+1} = x_{max}$; the other knots are arbitrary.