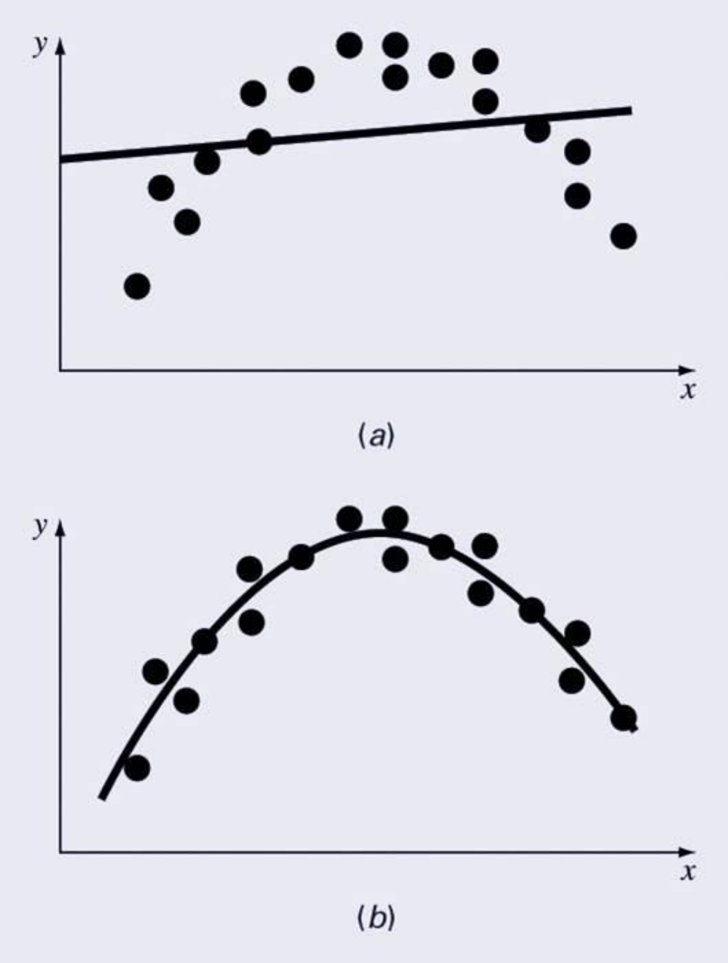
Chapter 15

General Linear Least-Squares and Nonlinear Regression

Numerical Methods
Fall 2019

Polynomial Regression

- The least-squares procedure from Chapter 13 can be extended to fit data to a higher-order polynomial.
- The idea is to minimize the sum of the squares of the estimate residuals.
- The figure shows the same data fit with:
  a) first order polynomial
  b) second order polynomial



Process and Measures of Fit

- For a second order polynomial, the best fit would mean minimizing:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

- In general, this would mean minimizing:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m)^2$$

- The standard error for fitting an m^{th} order polynomial to n data points is:

$$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$$

because the m^{th} order polynomial has $(m+1)$ coefficients.

- The coefficient of determination r^2 is still found using:

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

Process and Measures of Fit

- For a second order polynomial

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum x_i \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum x_i^2 \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

equations are set equal to zero

$$(n)a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i y_i$$

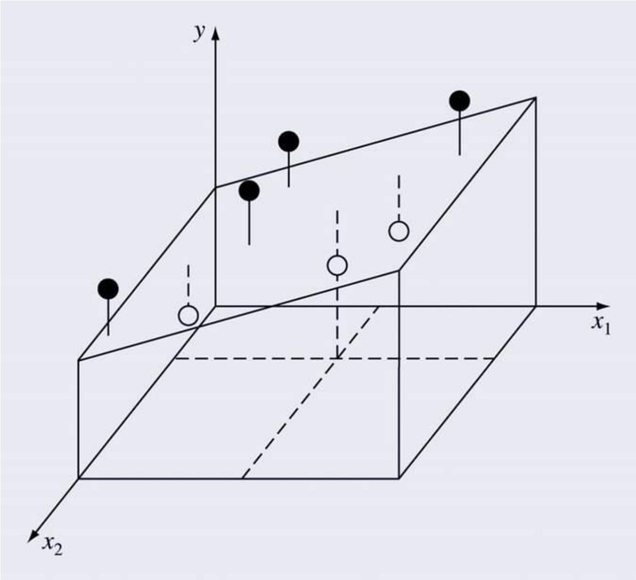
$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

EXAMPLE 15.1

Multiple Linear Regression

- Another useful extension of linear regression is the case where y is a linear function of two or more independent variables:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots a_m x_m$$



- Again, the best fit is obtained by minimizing the sum of the squares of the estimate residuals:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1,\,i} - a_2 x_{2,\,i} - \cdots a_m x_{m,\,i})^2$$

Multiple Linear Regression

- For example, y might be a linear function of x_1 and x_2,

$$y = a_0 + a_1 x_1 + a_2 x_2 + e$$

- The best values of the coefficients are determined by formulating the sum of the squares of the residuals

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i})^2$$

- and differentiating with respect to each of the unknown coefficients:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i})$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum x_{1,i} (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i})$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum x_{2,i} (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i})$$

Multiple Linear Regression

- The coefficients yielding the minimum sum of the squares of the residuals are obtained by setting the partial derivatives equal to zero and expressing the result in matrix form as

$$\begin{bmatrix} n & \sum x_{1,i} & \sum x_{2,i} \\ \sum x_{1,i} & \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{2,i} & \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1,i} y_i \\ \sum x_{2,i} y_i \end{Bmatrix}$$

EXAMPLE 15.2

General Linear Least Squares

- Linear, polynomial, and multiple linear regression all belong to the general linear least-squares model:

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots a_m z_m + e$$

where z_0, z_1, …, z_m are a set of m + 1 ***basis functions*** and e is the error of the fit.

- The basis functions can be any function of the data but ***cannot*** contain any of the coefficients a_0, a_1, etc.

Solving General Linear Least Squares Coefficients, 1

- The equation:

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots a_m z_m + e$$

can be re-written for each data point as a matrix equation:

$$\{y\} = [Z]\{a\} + \{e\}$$

where $\{y\}$ contains the dependent data, $\{a\}$ contains the coefficients of the equation, $\{e\}$ contains the error at each point, and $[Z]$ is:

$$[Z] = \begin{bmatrix} z_{01} & z_{11} & \cdots & z_{m1} \\ z_{02} & z_{12} & \cdots & z_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{0n} & z_{1n} & \cdots & z_{mn} \end{bmatrix}$$

with z_{ji} representing the value of the j th basis function calculated at the i th point.

Solving General Linear Least Squares Coefficients, 2

- Generally, [Z] is not a square matrix, so simple inversion cannot be used to solve for {a}. Instead the sum of the squares of the estimate residuals is minimized:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{m} a_j z_{ji} \right)^2$$

- The outcome of this minimization yields:

$$\left[[Z]^T [Z] \right] \{a\} = \{ [Z]^T \{y\} \}$$

MATLAB Example

- Given x and y data in columns, solve for the coefficients of the best fit line for $y=a_0+a_1x+a_2x^2$

```
Z = [ones(size(x) x x.^2]
a = (Z'*Z)\(Z'*y)
```

  - Note also that MATLAB's left-divide will automatically include the $[Z]^T$ terms if the matrix is not square, so

```
a = Z\y
```

  would work as well

- To calculate measures of fit:

```
St = sum((y-mean(y)).^2)
Sr = sum((y-Z*a).^2)
r2 = 1-Sr/St
syx = sqrt(Sr/(length(x)-length(a)))
```

EXAMPLE 15.3

Nonlinear Regression

- There are many cases in engineering and science where nonlinear models must be fit to data.
- For example,

$$y = a_0(1 - e^{-a_1 x})$$

- One method is to perform nonlinear regression to directly determine the least-squares fit.

$$f(a_0, a_1) = \sum_{i=1}^{n} [y_i - a_0(1 - e^{-a_1 x_i})]^2$$

Use MATLAB 'fminsearch' function

Nonlinear Regression in MATLAB

- To perform nonlinear regression in MATLAB, write a function that returns the sum of the squares of the estimate residuals for a fit and then use MATLAB's `fminsearch` function to find the values of the coefficients where a minimum occurs.

- The arguments to the function to compute S_r should be the coefficients, the independent variables, and the dependent variables.

Nonlinear Regression in MATLAB Example

- Given dependent force data F for independent velocity data v, determine the coefficients for the fit:

$$F = a_0 v^{a_1}$$

Example 14.6

- First, write a function called `fSSR.m` containing the following:

```
function f = fSSR(a, xm, ym)
yp = a(1)*xm.^a(2);
f = sum((ym-yp).^2);
```

- Then, use **`fminsearch`** in the command window to obtain the values of a that minimize **`fSSR`**:

```
a = fminsearch(@fSSR, [1, 1], [], v, F)
```

- where **`[1, 1]`** is an initial guess for the **`[a0, a1]`** vector, **`[]`** is a placeholder for the options

EXAMPLE 15.5

Nonlinear Regression Results

- The resulting coefficients will produce the largest r^2 for the data and may be different from the coefficients produced by a transformation: