Multiple Linear Regression

When doing a linear regression we attempt to find out if the expected value of some variable Y, which we will call dependent variable, can be modeled as a linear function of a series of other variables X_j $(j = 1, 2, \cdots, p)$, which we call independent variables:

$$E(Y|X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p. \tag{1}$$

Furthermore we might specify the variation of Y around its mean value by using the model

$$Y_i = \alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_p X_{p,i} + \epsilon_i = E(Y|X_i) + \epsilon_i, \tag{2}$$

where ϵ_i is a Gaussian error term with mean 0 and $X_{j,i}$ is the jth coefficient (or predictor) for the ith observation $(i = 1, 2, \cdots, n)$. The model can be written in matrix for as

$$\boldsymbol{Y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

or

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \tag{4}$$

The least squares problem consists now in minimizing the sum of the squares of the difference between the mean value of Y and its realization values Y_i. That is, looking for the coefficients $\hat{\boldsymbol{\beta}}$ that minimize

$$F = \sum_{i=1}^{n} \left(Y_i - \alpha - \sum_{j=1}^{p} \beta_j X_{i,j} \right)^2, \tag{5}$$

which has the solution

$$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^T \mathrm{X}\right)^{-1} \mathrm{X}^T \boldsymbol{Y}. \tag{6}$$

The fitted values have then the form

$$\begin{aligned} \hat{\boldsymbol{Y}} &= \mathrm{X}\hat{\boldsymbol{\beta}} \\ &= \mathrm{X}\left(\mathrm{X}^T \mathrm{X}\right)^{-1} \mathrm{X}^T \boldsymbol{Y} \\ &= \mathrm{H}\,\boldsymbol{Y}. \end{aligned} \tag{7}$$

0.1 Coefficient of multiple determination

The coefficient of multiple determination R^2 is defined as the ratio of the sum of the squared difference between the fitted and the average values

$$\begin{aligned} SSR =& \sum_{i=1}^{n} \left(\hat{Y}_i - \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2 \\ =& \boldsymbol{Y}^T \left[\mathrm{H} - \frac{1}{n} \mathrm{J} \right] \boldsymbol{Y} \end{aligned} \tag{8}$$

where J is an $n \times n$ matrix of ones, and the sum of the squared difference between the actual and average values (proportional to the variance)

$$\begin{aligned} SST =& \sum_{i=1}^{n} \left(Y_i - \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2 \\ =& \boldsymbol{Y}^T \left[\mathrm{I} - \frac{1}{n} \mathrm{J} \right] \boldsymbol{Y}, \end{aligned} \tag{9}$$

where I is the $n \times n$ identity matrix. That is

$$R^2 = \frac{SSR}{SST}. \tag{10}$$

The estimated variance is

$$\hat{\sigma}^2 = \frac{SST}{n - p - 1}. \tag{11}$$

0.2 Maximum likelihood

In our model we have assumed that the random variable Y has a Gaussian distribution around its mean and, assuming that the events are independent, the joint probability distribution for the values of $\boldsymbol{Y}$ obtained in the experiments given their mean and variance is

$$f(\boldsymbol{Y}|\mathrm{X}\boldsymbol{\beta}, \sigma^2 \mathrm{I}) = \left(\frac{1}{2\pi\sigma^2} \right)^{-n/2} e^{-\frac{1}{2\sigma^2} (\boldsymbol{Y} - \mathrm{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \mathrm{X}\boldsymbol{\beta})}, \tag{12}$$

and the logarithm of this (called the log likelihood) is

$$\ln[f(\boldsymbol{Y}|\mathrm{X}\boldsymbol{\beta}, \sigma^2 \mathrm{I})] = -\frac{1}{2\sigma^2} (\boldsymbol{Y} - \mathrm{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \mathrm{X}\boldsymbol{\beta}) + c \tag{13}$$

where c does not depend on $\boldsymbol{\beta}$. The maximum likelihood estimate for the parameters $\tilde{\boldsymbol{\beta}}$ is the one that maximizes eq (13). After differentiating and equating to 0 we get that

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \left(\mathrm{X}^T \mathrm{X} \right)^{-1} \mathrm{X}^T \boldsymbol{Y} \\ &= \hat{\boldsymbol{\beta}}. \end{aligned} \tag{14}$$