LINEAR REGRESSION AND POLYNOMIAL REGRESSION

SIEYE RYU

1. Simple Linear Regression

Simple linear regression is a linear regression of a data which has one independent variable. Suppose that we have n pairs of data

$$\{(x_i, y_i) \in \mathbb{R}^2 : i = 1, \cdots, n\}.$$

We will construct a linear predictor function $y = \beta_1 x + \beta_0$ from the data, based on the method of least squares.

The *residual* r_i $(i = 1, \cdots, n)$ is the difference between an observed value and the fitted value provided by a linear prediction:

$$r_i = y_i - (\beta_1 x_i + \beta_0) \qquad (i = 1, \cdots, n).$$

The method of least squares provides an approximation of the data by minimizing the sum of the squares of the residuals. In other words, if S is a function on $\mathbb{R}^2$ defined by

$$S(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_1 x_i - b_0)^2,$$

and if the point (β_0, β_1) minimizes $S(b_0, b_1)$, then the linear predictor function is given by

$$y = \beta_1 x + \beta_0.$$

Intuitively, $S(b_0, b_1)$ has a global minimum and it has no local maximum or saddle points. Hence, it is enough to find a critical point of S. To do this, we compute the gradient of S. By chain rule, we have

$$\nabla S(b_0, b_1) = \left(-2 \sum_{i=1}^{n} (y_i - b_1 x_i - b_0),\ -2 \sum_{i=1}^{n} x_i (y_i - b_1 x_i - b_0) \right).$$

If (β_0, β_1) is a critical point of S, then we have

$$\sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0) = 0 \tag{1.1}$$

$$\sum_{i=1}^{n} x_i (y_i - \beta_1 x_i - \beta_0) = 0. \tag{1.2}$$

The first equation (1.1) can be written as follows:

$$\sum_{i=1}^{n} (y_i - \beta_1 x_i) - n\beta_0 = 0.$$

This leads us to

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_1 x_i). \tag{1.3}$$

Substituting (1.3) into (1.2) yields

$$\sum_{i=1}^{n} x_i(y_i - \beta_1 x_i - \frac{1}{n}\sum_{j=1}^{n}(y_j - \beta_1 x_j)) = 0.$$

This is equivalent to

$$\beta_1\left(-\sum_{i=1}^{n} x_i^2 + \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) + \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 0.$$

Thus, we have

$$\beta_1 = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}. \tag{1.4}$$

We denote the means of x_i and y_i ($i = 1, \cdots, n$) by $\bar{x}$ and $\bar{y}$, respectively. Then (1.3) is equivalent to

$$\beta_0 = \bar{y} - \beta_1 \bar{x}.$$

In the rest of the section, we prove that (1.4) is equivalent to

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \tag{1.5}$$

The following calculation shows that the numerator of (1.5) is the same as that of (1.4):

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right) \\
&= \sum_{i=1}^{n} x_i y_i - \frac{2}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) + \frac{n}{n^2}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \\
&= \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right).
\end{aligned}$$

Replacing y_i and $\bar{y}$ with x_i and $\bar{x}$ shows that the denominators of (1.4) and (1.5) are the same.

2. Multivariate Polynomial Regression

Multivariate polynomical regression is a polynomial regression for a data having more than one independent variables. In this section, we only consider the case where a data has two independent variables. Suppose that we have n triples of data

$$\{(p_i, q_i, y_i) \in \mathbb{R}^2 : i = 1, \cdots, n\}.$$

We will establish a polynomial regression of degree 2,

$$y = \beta_0 + \beta_1 p + \beta_2 q + \beta_3 p^2 + \beta_4 pq + \beta_5 q^2$$

from the data, based on the method of least squares.

The residual r_i is given by

$$r_i = y_i - (\beta_0 + \beta_1 p_i + \beta_2 q_i + \beta_3 p_i^2 + \beta_4 p_i q_i + \beta_5 q_i^2) \qquad (i = 1, \cdots, n).$$

The sum of the squares of the residuals is given by

$$S = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 p_i + \beta_2 q_i + \beta_3 p_i^2 + \beta_4 p_i q_i + \beta_5 q_i^2) \right]^2.$$

We regard S as a function of $b_0, b_1, \cdots, b_5$ and find the point $(\beta_0, \cdots, \beta_5)$ at which $S(b_0, \cdots, b_5)$ has the global minimum. Intuitively, $S(b_0, \cdots, b_5)$ has a global minimum and it has no local maximum or saddle points. Hence it suffices to find a critical point of S. To do this, we compute the gradient of S at the critical point $(\beta_0, \cdots, \beta_5)$. By chain rule, we have

$$\nabla S = -2 \sum_{i=1}^{n} (r_i,\, p_i r_i,\, q_i r_i,\, 2p_i r_i,\, p_i q_i r_i,\, 2q_i r_i).$$

Since $\nabla S(\beta_0, \cdots, \beta_5) = 0$, it follows that

$$\sum_{i=1}^{n} r_i = 0, \qquad \sum_{i=1}^{n} p_i r_i = 0, \qquad \sum_{i=1}^{n} q_i r_i = 0,$$
$$\sum_{i=1}^{n} p_i r_i = 0, \qquad \sum_{i=1}^{n} p_i q_i r_i = 0, \qquad \sum_{i=1}^{n} q_i r_i = 0.$$

References

[1] Linear regression on Wikipedia
[2] Simple linear regression on Wikipedia
[3] Least squares on Wikipedia