Brian Tenneson

Discussion 3

ADS 534

7/16/2024

In the link provided by the professor (`https://online.stat.psu.edu/stat462/node/132/`) note that there are two parameters β_0 and β_1 but linear regression can be carried out with more parameters.

Let y be a column vector of length n. y is called the response variable and the dependent variable. Let ε also be a length n column vector. This represents the error term for each observation. β is a column vector of length $p+1$. The $p+1$ elements of β are the changes in the expected responses (i.e., $E(Y)$)) per unit change in the other variables in x. x is $n \times (p+1)$ and has the following form:

$$x = \begin{pmatrix} x_{11} & \dots & x_{1p} & 1 \\ \vdots & x_{ij} & \vdots & \vdots \\ x_{n1} & \dots & x_{np} & 1 \end{pmatrix}.$$

This leaves β_0 to be the intercept of the least squares regression.

With this setup in mind, we have

$$y = x\beta + \varepsilon.$$

This is why it is called linear regression instead of, for example, logistic regression.

x_{ij} represents the value of the j-th predictor variable for the i-th observation.

The last column of 1's in x corresponds to the intercept term in the regression model.

So, the x's are the values of the predictor variables used to predict the response variable y in the linear regression model.

The least squares solution $\beta := \begin{pmatrix} \beta_0 & \beta_1 & \dots & \beta_p \end{pmatrix}^t$ is given by

$$(x^t x)^{-1} x^t y.$$

Note that $x^t x$ is square with dimensions $(p+1) \times (p+1)$ and if the columns of $x^t x$ are linearly independent, the invertibility of the matrix $x^t x$ is assured. x^t has dimensions $(p+1) \times n$ and so $x^t y$ has dimensions $(p+1) \times 1$. Thus $(x^t x)^{-1} x^t y$ has dimensions $(p+1) \times 1$ which corresponds to our solution β which is of size $(p+1) \times 1$.

What does it mean about the data that might ensure that the columns of $x^t x$ are linearly independent to ensure the existence of a least squares solution?

The following quote is taken from chatgpt4:

1. **Non-collinearity of Predictors**:

For the columns of $x^T x$ to be linearly independent, the predictor variables represented by the columns of x (excluding the last column of ones for the intercept) must not be perfectly collinear. This means no predictor can be a perfect linear combination of others. If one predictor is a linear combination of others, then the matrix $x^T x$ will have linearly dependent columns, leading to multicollinearity, which makes $x^T x$ non-invertible.

2. **Sufficient Observations**:

The matrix x needs to have a sufficient number of observations (rows) relative to the number of predictors (columns, excluding the intercept) to avoid a situation where $x^T x$ is singular. Typically, you want more observations than predictors ($n > p$). If there are fewer observations than predictors, or they are equal, $x^T x$ can be singular or not full rank, meaning it does not cover all dimensions (it is not invertible).

3. **Variability in Data**:

The data in each predictor variable must vary; if a predictor variable has little or no variation (it is constant), it does not help in predicting the response variable, and its inclusion can lead to a rank-deficient $x^T x$ matrix. For instance, if one of the predictors were a constant (the same value for all observations), this would reduce the rank of x, thereby affecting the invertibility of $x^T x$.

4. **Avoidance of Exact Multicollinearity**:

Exact multicollinearity is a scenario where one predictor is an exact linear combination of one or more of the other predictors. This must be avoided because it directly leads to dependent columns in x, and hence in $x^T x$, making $x^T x$ singular (and therefore not invertible).

5. **Handling of Outliers**:

Outliers can influence the estimates of β significantly. While outliers do not directly impact the invertibility of $x^T x$, they can affect the solution's stability and accuracy. Robust regression methods might be necessary if outliers are present.

In summary, ensuring that the columns of $x^T x$ are linearly independent generally involves having a sufficiently large and varied dataset, with each predictor providing unique information that is not a simple reiteration of what other predictors are providing. This ensures a stable and meaningful solution to the least squares problem in linear regression.