Regression Analysis
Multiple Linear Regression

Buddhananda Banerjee

Department of Mathematics
Centre for Excellence in Artificial Intelligence
Indian Institute of Technology Kharagpur

bbanerjee@maths.iitkgp.ac.in

Multiple linear regression

- Consider a data set $D = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^{k+1}, y_i \in \mathbb{R}, \ \forall i = 1, 2, \cdots, n\}$
- $\mathbf{x}_i$s are non stochastic
- y_is are stochastic and realized values of random variable Y_is
- ϵ_is are iid $N(0, \sigma^2)$ random variables
- Regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)^T$ is unknown
- Error parameter σ^2 is unknown

Problem statement

Considering the linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \ \ \forall \ \ i = 1, 2, \ldots n$$

we want to estimate $\boldsymbol{\beta}$ which will minimize least squared condition.

Matrix representation

For n such observation we use matrix notation to represent it as follows,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where, $\mathbf{Y} = (y_1, y_2, \cdots y_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)^T$,
$\mathbf{X} = (\mathbf{1}, \mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_k) = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \cdots \epsilon_n)^T$.
Hence, there are $k+2$ unknown model parameters, $\boldsymbol{\beta}$ and $\sigma^2 > 0$

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \;\; as \;\; \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n) \tag{2}$$

Problem statement

The least square condition to be minimized to estimate $\boldsymbol{\beta}, \sigma^2$ is

$$S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}^T\mathbf{Y} - 2\beta^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \tag{3}$$

Least square

Estimation

If $\hat{\beta}$ minimizes the least square condition then it satisfies the normal equations

$$\begin{aligned} & \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0} \\ \implies \quad & -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \\ \implies \quad & \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} \qquad (4) \end{aligned}$$

Least square

Prediction

Fitted regression for the used data is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} = P_{\mathbf{X}}\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 P_{\mathbf{X}})$$

where $P_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ is the orthogonal projection matrix of the column space of $\mathbf{X}$ i.e. $\mathcal{C}(\mathbf{X})$. It means $\hat{\mathbf{y}} \in \mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{X}^T\mathbf{X})$.

Least square

Estimated Error

Hence the estimated error in prediction

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_n - P_{\mathbf{X}})\mathbf{y} \sim N(\mathbf{0}, \sigma^2(\mathbf{I}_n - P_{\mathbf{X}}))$$

where $\mathbf{e} \in \mathcal{C}(\mathbf{X})^{\perp} = \mathcal{C}(\mathbf{X}^T\mathbf{X})^{\perp}$.

Note: Hence $\hat{\mathbf{y}}$ and $\mathbf{e}$ are uncorrelated and they are independently distributed when $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Definition

Linear unbiased estimator (LUE)

A linear estimator $\mathbf{l}^T\mathbf{y} = \sum_i l_i y_i$ is said to be linear unbiased estimator (LUE) of $\mathbf{p}^T\beta$ if $E(\mathbf{l}^T\mathbf{y}) = \mathbf{p}^T\beta$ for all $\beta \in \mathbb{R}^{k+1}$.

Linear zero function (LZF)

A linear estimator $\mathbf{l}^T\mathbf{y} = \sum_i l_i y_i$ is said to be linear zero function (LZF) if $E(\mathbf{l}^T\mathbf{y}) = 0$ for all $\beta \in \mathbb{R}^{k+1}$

Definition

Estimablity

A linear parametric function $\mathbf{p}^T\beta$ is said to be **estimable** if it has a LUE i.e. there exists $\mathbf{l}^T\mathbf{y}$ such that $E(\mathbf{l}^T\mathbf{y}) = \mathbf{p}^T\beta$ for all $\beta \in \mathbb{R}^{k+1}$.

BLUE

The best linear unbiased estimator (BLUE) of a linear parametric function $p^T\beta$ is a LUE with minimum variance.

Best estimator

Theorem

A linear function is BLUE of its expectation iff it is uncorrelated with all linear zero functions (LZF).

Corollary

If $\mathbf{l}^T\mathbf{y}$ is an LUE of $\mathbf{p}^T\beta$ then $\mathbf{l}^T P_{\mathbf{X}}\mathbf{y}$ is the BLUE of $\mathbf{p}^T\beta$

Note: If you know a LUE of a parametric function then you can get the BLUE out of it.

ANOVA

To test $H_0 : \beta_R = (\beta_1, \cdots, \beta_k) = \mathbf{0}$ vs $H_1 : (\beta_1, \cdots, \beta_k) \neq \mathbf{0}$ we perform the ANOVA as

$$SST = SSModel + SSRes$$

- $SST = \mathbf{Y}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{Y} \sim \sigma^2\chi^2(df = n-1, ncp = \lambda/\sigma^2)$
- $SSModel = \mathbf{Y}^T(P_X - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{Y} \sim \sigma^2\chi^2(df = k, ncp = \lambda/\sigma^2)$
- $SSRes = \mathbf{Y}^T(\mathbf{I}_n - P_X)\mathbf{Y} \sim \sigma^2\chi^2(df = n-k-1, ncp = 0)$

Under $H_0 : \lambda = 0$, otherwise $\lambda = \beta_R^T \mathbf{X}_c^T \mathbf{X}_c \beta_R$ where $\mathbf{X}_c$ is the centred regressor variables for β_R. Hence, by Cochran's theorem we have

$$\frac{SSModel/k}{SSRes/(n-k-1)} \sim F_{k,(n-k-1),ncp=\lambda}$$

It is a right tailed test because $\lambda = 0$ under H_0.


THANK
YOU