Statistical Inference and Multivariate Analysis (MA324)

Lecture Slides
Lecture 29

Multiple Linear Regression

Indian Institute of Technology Guwahati

Jan-May 2023

Multiple Linear Regression

- In general, the response (y) may be related to p regressors (input variables/predictors).
- The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i, i = 1, \ldots, n$$

  is called **multiple linear regression**. A regression model with p regressors.
- The parameters $\beta_j, j = 0, 1, 2, \cdots, p$ are called regression coefficients.
- $\beta_j, j = 0, 1, 2, \cdots, p$ represents the change in the average value of the response for a unit change in j^{th} regressor keeping other regressors fixed.
- As before, ϵ_i's are i.i.d. with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$ for $i = 1, \ldots, n$

- Data (of sample size n): $(y_i, x_{i1}, x_{i2}, \cdots, x_{ip}), i = 1, \ldots, n$
- Scalar notation becomes cumbersome – so matrix notation is used
- Let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ \vdots & \vdots & x_{ij} & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix}, \; \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \; \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

- Then we can write the model in a more compact form:

$$y_{n\times 1} = X_{n\times(p+1)}\beta_{(p+1)\times 1} + \epsilon_{n\times 1}$$

- X is called the ***design matrix***

Multiple Linear Regression Model

- Multiple Linear Regression Model:

$$y = X\beta + \epsilon$$

- ϵ is a **random vector** rather than a random variable.
- Assumptions: here, $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I$ and all the assumptions stated in simple linear regression.
- Note that Var is an abuse of notation; in the present context it really means the **"variance-covariance matrix".**

Estimation of Model Parameters:

- The LSEs of $\beta_0, \beta_1, \cdots, \beta_p$ are

$$
\begin{aligned}
& \underset{\beta_0,\beta_1,\cdots,\beta_p}{\operatorname{argmin}} \; Q(\beta_0, \beta_1, \cdots, \beta_p) \\
= & \underset{\beta_0,\beta_1,\cdots,\beta_p}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_p x_{ip} \Big)^2 \\
= & \underset{\beta}{\operatorname{argmin}} \; (y - X\beta)^T (y - X\beta) \\
= & \underset{\beta}{\operatorname{argmin}} (y^T y - 2\beta^T X^T y + \beta^T X^T X \beta)
\end{aligned}
$$

- How to differentiate $Q(\beta)$ with respect to β?

Estimation of Model Parameters:

- **Differentiating** $Q(\beta)$ **with respect to** β and setting the **derivative to zero** gives the following ***normal equations*****:**

$$X^T X\beta = X^T y$$

- Now **if the matrix** $X^T X$ **is *invertible*** (i.e. if X is of *full rank*), then the LSE is given by

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Fitted Value:

- The fitted value of response corresponding to regression $\underset{\sim}{x} = (1, x_1, \cdots, x_p)$ is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p$$

- Then $\underset{\sim}{\hat{y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n)^T = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy$, where, $H = X(X^T X)^{-1} X^T$ is called **hat-matrix**.

- The residuals:

$$\underset{\sim}{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} = \underset{\sim}{y} - \underset{\sim}{\hat{y}} = (I - H)\underset{\sim}{y}$$

Properties of LSE of $\underset{\sim}{\beta}$: Theorems

- $\underset{\sim}{\hat{\beta}}$ is a linear function of $\underset{\sim}{y}$
- $\underset{\sim}{\hat{\beta}}$ is unbiased estimator of $\underset{\sim}{\beta}$. That is, $E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta}$
- $Var(\underset{\sim}{\hat{\beta}}) = \sigma^2(X^T X)^{-1}$
- $\underset{\sim}{\hat{\beta}}$ is the BLUE of $\underset{\sim}{\beta}$.