Multiple Linear Regression using Matrices - I I

- Where
  - for subject i, $\boldsymbol{x}_i^T = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$ is a vector of observations on k *covariates* (also known as *predictor variables*, or *independent variables*). The "1" is needed when an intercept is included in the regression model.
  - Y_i is the observation on the *outcome variable* (also known as the *dependent variable*)
  - $\beta^T = (\beta_0, \beta_1, \ldots, \beta_k)$ is a vector of constants
- the *multiple regression model* says that for the i'th subject

$$Y_i = \mu_i + \epsilon_i,$$

where

$$\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta},$$

or

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

Multiple Linear Regression using Matrices - I II

- We assume to begin that the ϵ_i are mutually uncorrelated, and have zero mean and constant variance σ^2.
- The model says that the observation on the i'th subject consists of its mean, which is a linear function of the covariates, and an additive error term ϵ_i.

- Collecting all terms into vectors and matrices gives

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,k} \\ 1 & x_{21} & x_{22} & \dots & x_{2,k} \\ 1 & x_{31} & x_{32} & \dots & x_{3,k} \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 1 & x_{n1} & x_{n2} & \dots & x_{n,k} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix},$$

- or

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Least squares estimation

- The method of least squares is usually used for estimating $\boldsymbol{\beta}$, that is we find $\hat{\boldsymbol{\beta}}$ which minimizes

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

- The least squares estimates can be found by setting the vector of partial derivatives of $S(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ equal to 0.
- Taking derivatives gives (see Appendix C.2.2)

$$-2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

- Setting to zero gives the 'normal equations'

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0,$$

which are solved to give

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