

# Multiple Linear Regression using Matrices - I I

- Where
  - for subject  $i$ ,  $\mathbf{x}_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ik})$  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*)
  - $\beta^T = (\beta_0, \beta_1, \dots, \beta_k)$  is a vector of constants
- the *multiple regression model* says that for the  $i$ 'th subject

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

where

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

or

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

# Multiple Linear Regression using Matrices - I II

- We assume to begin that the  $\epsilon_i$  are mutually uncorrelated, and have zero mean and constant variance  $\sigma^2$ .
- 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  $\epsilon_i$ .

- Collecting all terms into vectors and matrices gives

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

- or

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

# Least squares estimation

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

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

- The least squares estimates can be found by setting the vector of partial derivatives of  $S(\beta)$  with respect to  $\beta$  equal to 0.
- Taking derivatives gives (see book Appendix C.2.2, or optional notes on differentiating wrt a vector)

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

- Setting to zero gives the 'normal equations'

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

which are solved to give

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