

Multiple Linear Regression

- ▶ **features/predictors:** X_1, \dots, X_p
- ▶ **response/outcome** variable: Y

The linear regression model assumes

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

where

β_0 is the intercept

β_j is the regression coefficient associated with X_j

e is the error term often assumed to have mean zero and variance σ^2 .

Multiple Linear Regression

- ▶ **features/predictors:** X_1, \dots, X_p
- ▶ **response/outcome** variable: Y

The linear regression model assumes

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

where

β_0 is the intercept

β_j is the regression coefficient associated with X_j

e is the error term often assumed to have mean zero and variance σ^2 .

Multiple Linear Regression

- ▶ **features/predictors:** X_1, \dots, X_p
- ▶ **response/outcome** variable: Y

The linear regression model assumes

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

where

β_0 is the intercept

β_j is the regression coefficient associated with X_j

e is the error term often assumed to have **mean zero**
and **variance σ^2** .

Multiple Linear Regression

- ▶ **features/predictors:** X_1, \dots, X_p
- ▶ **response/outcome** variable: Y

The linear regression model assumes

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

where

β_0 is the intercept

β_j is the regression coefficient associated with X_j

e is the error term often assumed to have mean zero and variance σ^2 .

Housing Data

Y : sale price of a house

X_1 : # of bedrooms

X_2 : # of bathrooms

X_3 : square feet

.....

Multiple Linear Regression

- ▶ **features/predictors:** X_1, \dots, X_p
- ▶ **response/outcome** variable: Y

The linear regression model assumes

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

where

β_0 is the intercept

β_j is the regression coefficient associated with X_j

e is the error term often assumed to have mean zero and variance σ^2 .

Housing Data

Y : sale price of a house

X_1 : # of bedrooms

X_2 : # of bathrooms

X_3 : square feet

.....

Training Data $(x_{i1}, \dots, x_{ip}, y_i)_{i=1}^n$

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i$$

$$i = 1, \dots, n$$

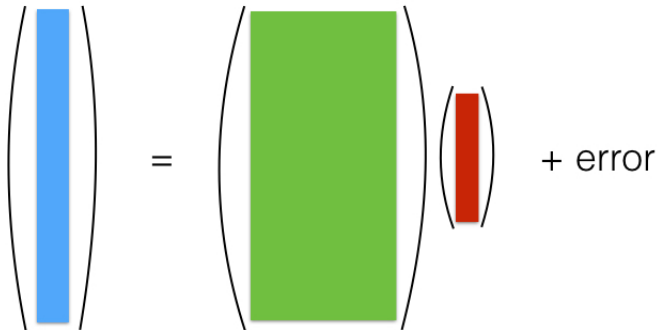
Matrix Representation

Express the regression model on $(x_{i1}, \dots, x_{ip}, y_i)_{i=1}^n$ in the following matrix form

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \dots + x_{1p}\beta_p + e_1 \\ \beta_0 + x_{21}\beta_1 + x_{22}\beta_2 + \dots + x_{2p}\beta_p + e_2 \\ \dots \\ \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 + \dots + x_{np}\beta_p + e_n \end{pmatrix}$$
$$= \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ 1 & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

$$\mathbf{y}_{n \times 1} = \mathbf{X}_{n \times (p+1)} \boldsymbol{\beta}_{(p+1) \times 1} + \mathbf{e}_{n \times 1}$$

The classical **large n small p** regression model:



Focus of **this** week

The modern **large p small n** regression model:



The diagram illustrates the regression model equation $y = X\beta + \epsilon$ using colored shapes and vector notation. On the left, a blue vertical rectangle is enclosed in large parentheses, representing the response vector y . This is followed by an equals sign. In the center, a large green square is enclosed in large parentheses, representing the design matrix X . To the right of the green square is a red vertical rectangle, also enclosed in large parentheses, representing the coefficient vector β . Finally, the text "+ error" is placed to the right of the red vector, representing the error term ϵ .

Focus of **next** week