Contents

Matrix Notation for Simple Linear Regression (Ch.5)

Simple Linear Regression Model

Recall the simple Linear Regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where ε_i's are independent Normally distributed random variables with mean 0 and variance σ^2.

Consider writing the observations:

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

The above equation group can be written in matrix form:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
$$= \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

- $Y_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}$ be the vector of responses.

- $X_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n\times 2}$ be the design matrix.

- $\beta_{2\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}_{2\times 1}$ be the vector of parameters.

- $\varepsilon_{n\times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$ be the vector of error terms is:

The simple linear regression model can then be written as:

$$Y_{n\times 1} = X_{n\times 2}\beta_{2\times 1} + \varepsilon_{n\times 1}$$
$$Y = X\beta + \varepsilon$$

Estimations

$$\hat{\beta}_{2\times 1} = b_{2\times 1} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}_{2\times 1} = (X_{n\times 2}^T X_{n\times 2})^{-1}(X_{n\times 2}^T Y_{n\times 2}) = (X^T X)^{-1}(X^T Y),$$

and

$$\mathrm{cov}(b_{2\times 1}) = \sigma^2 (X^T X)^{-1}.$$

The variance σ^2 can be estimated by MSE. $\widehat{\sigma^2} = MSE$.

Prediction

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = Xb$$

To predict Y at X_h, $\hat{Y}_h = \begin{pmatrix} 1 & X_h \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$.

The Hat Matrix

Note that $\hat{Y} = Xb$, and $b = \left(X^T X\right)^{-1}\left(X^T Y\right)$. Hence,

$$\hat{Y} = Xb = X\left(X^T X\right)^{-1}\left(X^T Y\right) = \left(X\left(X^T X\right)^{-1} X^T\right)Y = HY$$

where $H = X\left(X^T X\right)^{-1} X^T$ is called the "**hat matrix**."

Partitioning Sum of Squares: ANOVA

Let $I_{n\times n}$ be a diagonal matrix with 1's on the diagonal and 0 everywhere else (i.e., Identity matrix). Let $J_{n\times n}$ be a matrix of 1's . Then:

- The sum of squares explained by the regression: $SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = Y'\left(H - \frac{1}{n}J\right)Y$
- The sum of squares of error: $SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 = Y'\left(I - H\right)Y$
- Total sum of squares: $SSTO = \sum\left(Y_i - \bar{Y}\right)^2 = Y'\left(I - \frac{1}{n}J\right)Y$

- df_R = 1 = rank $\left(H - \frac{1}{n}J\right)$
- df_E = n – 2 = rank $\left(I - H\right)$
- df_{To} = n – 1 = rank $\left(I - \frac{1}{n}J\right)$

Recall that: SSR + SSE = SSTO, $df_R + df_E = df_{To}$, and MSR = SSR/df_R, MSE = SSE/df_E.

Multiple Linear Regression (Ch. 6)

The work is very similar to simple linear regression.

Multiple Linear Regression Model

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{1,2} + \cdots + \beta_{p-1} X_{1,p-1} \\ \beta_0 + \beta_1 X_{2,1} + \beta_2 X_{2,2} + \cdots + \beta_{p-1} X_{2,p-1} \\ \vdots \\ \beta_0 + \beta_1 X_{n,1} + \beta_2 X_{n,2} + \cdots + \beta_{p-1} X_{n,p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$= \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & & \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

- $X_{n\times p} = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & & \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{pmatrix}$ be the design matrix.
- $\beta_{p\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$ be the vector of parameters.

The multiple linear regression model can then be written as:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$$
$$Y = X\beta + \varepsilon$$

Estimations

$$\hat{\beta}_{p\times 1} = b_{p\times 1} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_{p-1} \end{pmatrix}_{p\times 1} = \left(X^T_{n\times p}X_{n\times p}\right)^{-1}\left(X^T_{n\times p}Y_{n\times p}\right) = (X^TX)^{-1}(X^TY),$$

and

$$cov(\, b) = \sigma^2(X^TX)^{-1}.$$

The variance σ^2 can be estimated by MSE. $\widehat{\sigma^2} = MSE$

Prediction

- The fitted values (for observations in the data set):

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_{1,1} + b_1 X_{1,2} + \cdots + b_{p-1} X_{1,p-1} \\ b_0 + b_1 X_{2,1} + b_2 X_{2,2} + \cdots + b_{p-1} X_{2,p-1} \\ \vdots \\ b_0 + b_1 X_{n,1} + b_2 X_{n,1} + \cdots + b_{p-1} X_{n,p-1} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & & \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p-1} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_{p-1} \end{pmatrix} = Xb.$$

- The predicted value (aka. the esimate of Y, or the fitted value) for a new set of x-values: $(X_{h,1} \quad X_{h,2} \quad \cdots \quad X_{h,p-1})$.

$$\hat{Y}_h = (1 \quad X_{h,1} \quad X_{h,2} \quad \cdots \quad X_{h,p-1}) b.$$

Partitioning Sum of Squares: ANOVA

Let $H = X\left(X^T X\right)^{-1} X^T$ denote the hat matrix. Let $I_{n\times n}$ be a diagonal matrix with 1's on the diagonal and 0 everywhere else (i.e., Identity matrix). Let $J_{n\times n}$ be a matrix of 1's . Then:

- The sum of squares explained by the regression: $SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = Y'\left(H - \frac{1}{n}J\right)Y$
- The sum of squares of error: $SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 = Y'\left(I - H\right)Y$
- Total sum of squares: $SSTO = \sum\left(Y_i - \bar{Y}\right)^2 = Y'\left(I - \frac{1}{n}J\right)Y$

- df_R = $p - 1$ = rank $\left(H - \frac{1}{n}J\right)$
- df_E = $n - p$ = rank $\left(I - H\right)$
- df_{TO} = $n - 1$ = rank $\left(I - \frac{1}{n}J\right)$

Recall that: SSR + SSE = SSTO, $df_R + df_E = df_{TO}$, and MSR = SSR/df_R, MSE = SSE/df_E.