STAT 525

Chapter 5
Matrix Approach to Simple Linear Regression

Dr. Qifan Song

Matrix

- Collection of elements arranged in rows and columns
- Elements will be numbers or symbols
- For example:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 2 & 6 \end{bmatrix}$$

- Rows denoted with the i subscript
- Columns denoted with the j subscript
- The element in row 1 col 2 is 3
- The element in row 3 col 1 is 2

- Elements often expressed using symbols

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1c} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & a_{r3} & \cdots & a_{rc} \end{bmatrix}$$

- Matrix $\mathbf{A}$ has r rows and c columns

- Said to be of dimension $r \times c$

- Element a_{ij} is in i^{th} row and j^{th} col

- A matrix is square if $r = c$

- Called a column vector if $c = 1$

- Called a row vector if $r = 1$

Matrix Operations

- Transpose
  - Denoted as $\mathbf{A}'$

    Row 1 becomes Column 1, Row r becomes Column r

    $$\downarrow$$

    Column 1 becomes Row 1, Column c becomes Row c
  - If $\mathbf{A} = [a_{ij}]$ then $\mathbf{A}' = [a_{ji}]$
  - If $\mathbf{A}$ is $r \times c$ then $\mathbf{A}'$ is $c \times r$

- Addition and Subtraction
  - Matrices must have the same dimension
  - Addition/subtraction done on element by element basis

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1c} + b_{1c} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} + b_{r1} & a_{r2} + b_{r2} & \cdots & a_{rc} + b_{rc} \end{bmatrix}$$

- Multiplication
  - If scalar then $\lambda \mathbf{A} = [\lambda a_{ij}]$
  - If multiplying two matrices $(\mathbf{C} = \mathbf{AB})$
    * $c_{ij} = \sum_k a_{ik} b_{kj}$
    * Columns of $\mathbf{A}$ must equal Rows of $\mathbf{B}$
    * Resulting matrix of dimension Rows(A) $\times$ Columns(B)
  - Elements obtained by taking cross products of rows of $\mathbf{A}$ with columns of $\mathbf{B}$

$$\begin{bmatrix} 1 & 2 \\ 4 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 4 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 3 \\ 17 & 10 & 5 \\ 15 & 12 & 6 \end{bmatrix}$$

Regression Matrices

- Consider example with $n = 4$
- Consider expressing observations:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 &+\varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 &+\varepsilon_2 \\ Y_3 &= \beta_0 + \beta_1 X_3 &+\varepsilon_3 \\ Y_4 &= \beta_0 + \beta_1 X_4 &+\varepsilon_4 \end{aligned}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \beta_0 + \beta_1 X_3 \\ \beta_0 + \beta_1 X_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\mathbf{X}$ is called the design matrix

Special Regression Examples

- Using multiplication and transpose

$$
\begin{aligned}
\mathbf{Y}'\mathbf{Y} &= \sum Y_i^2 \\
\mathbf{X}'\mathbf{X} &= \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \\
\mathbf{X}'\mathbf{Y} &= \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}
\end{aligned}
$$

- Will use these to compute $\widehat{\boldsymbol{\beta}}$ etc.

Special Types of Matrices

- Symmetric matrix
  - When $\mathbf{A} = \mathbf{A}'$
  - Requires $\mathbf{A}$ to be square
  - Example: $\mathbf{X}'\mathbf{X}$

- Diagonal matrix
  - Square matrix with off-diagonals equal to zero
  - Important example: Identity matrix

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

  - $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$

Linear Dependence

- Consider the matrix

$$\mathbf{Q} = \begin{bmatrix} 5 & 3 & 10 \\ 1 & 2 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

the columns of $\mathbf{Q}$ are vectors.

$$\mathbf{C}_1 = \begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{C}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{C}_3 = \begin{bmatrix} 10 \\ 2 \\ 2 \end{bmatrix}$$

- If there *is* a relationship between the columns of a matrix such that

$$\lambda_1 \mathbf{C}_1 + \ldots + \lambda_c \mathbf{C}_c = \mathbf{0}$$

and not all λ_j's are 0, then the set of column vectors are *linearly dependent*.

  - For the above example, $-2\mathbf{C}_1 + 0\mathbf{C}_2 + 1\mathbf{C}_3 = 0$.

- If such a relationship does not exist then the set of columns are *linearly independent.*

  - Columns of an identity matrix are linearly indpendent.

- Similarly consider rows

Rank of a Matrix

- The **rank** of a matrix is the maximum number of linear independent columns (or rows)
- Rank of a matrix cannot exceed min(r, c)
- **Full Rank** $\equiv$ all columns are linearly independent
- Example:

$$\mathbf{Q} = \begin{bmatrix} 5 & 3 & 10 \\ 1 & 2 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

  - The rank of $\mathbf{Q}$ is 2
- Rank of matrix can be connected to the d.f.

Inverse of a Matrix

- Inverse similar to the reciprocal of a scalar
- Inverse defined for square matrix of full rank
- Want to find the inverse of **S**, such that

$$\mathbf{S} \cdot \mathbf{S}^{-1} = \mathbf{I}$$

- Easy example: Diagonal matrix
  - Let $\mathbf{S} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ then

$$\mathbf{S}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

inverse of each element on the diagonal

- General procedure for 2×2 matrix
- Consider:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

1. Calculate the *determinant* $D = a \cdot d - b \cdot c$

   If $D = 0$ then the matrix has no inverse.

2. In $\mathbf{A}^{-1}$, switch a and d; make c and b negative; multiply each element by $\frac{1}{D}$

$$\mathbf{A}^{-1} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{bmatrix}$$

   - Steps work only for a 2×2 matrix.
   - Algorithm for 3×3 given in book

Use of Inverse

- Consider equation $2x = 3 \longrightarrow x = 3 \times \frac{1}{2}$
- Inverse similar to using reciprocal of a scalar
- Pertains to a set of equations

$$\begin{array}{ccc} \mathbf{A} & \mathbf{X} = & \mathbf{C} \\ (r \times r) & (r \times 1) & (r \times 1) \end{array}$$

- Assuming A has an inverse:

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{A}\mathbf{X} &= \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{X} &= \mathbf{A}^{-1}\mathbf{C} \end{aligned}$$

Random Vectors and Matrices

- Contain elements that are random variables
- Can compute expectation and (co)variance
- In regression set up, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, both $\boldsymbol{\varepsilon}$ and $\mathbf{Y}$ are random vectors
- Expectation vector: $E(\mathbf{Y}) = [E(Y_i)]$
- Covariance matrix: symmetric

$$\boldsymbol{\sigma}^2(\mathbf{Y}) = \begin{bmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \cdots & \sigma(Y_1, Y_n) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \cdots & \sigma(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma(Y_n, Y_1) & \sigma(Y_n, Y_2) & \cdots & \sigma^2(Y_n) \end{bmatrix}$$

Basic Theorems

- Consider random vector $\mathbf{Y}$
- Consider constant matrix $\mathbf{A}$
- Suppose $\mathbf{W} = \mathbf{AY}$
  - $\mathbf{W}$ is also a random vector
  - $E(\mathbf{W}) = \mathbf{A} \times E(\mathbf{Y})$
  - $\boldsymbol{\sigma}^2(\mathbf{W}) = \mathbf{A} \times \boldsymbol{\sigma}^2(\mathbf{Y}) \times \mathbf{A}'$
- If $\mathbf{Y}$ is a multivariate normal, then $\mathbf{W} = \mathbf{AY}$ is multivariate normal as well.

Regression Matrices

- Can express observations

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Both $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are random vectors

$$\begin{aligned} E(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} + E(\boldsymbol{\varepsilon}) \\ &= \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\sigma}^2(\mathbf{Y}) &= 0 + \boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}) \\ &= \sigma^2\mathbf{I} \end{aligned}$$

Least Squares

- Express quantity Q

$$\begin{aligned} Q &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

  - $(\mathbf{X}\boldsymbol{\beta})' = \boldsymbol{\beta}'\mathbf{X}'$

- Taking derivative $\longrightarrow -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0$

  - $\frac{\partial}{\partial \beta}\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{Y}$

  - $\frac{\partial}{\partial \beta}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$

- This means $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Fitted Values

- The fitted values $\hat{\mathbf{Y}} = \mathbf{X}b = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- Matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the *hat matrix*
  - $\mathbf{H}$ is symmetric, i.e., $\mathbf{H}' = \mathbf{H}$
  - $\mathbf{H}$ is idempotent, i.e., $\mathbf{HH} = \mathbf{H}$
- Equivalently write $\hat{\mathbf{Y}} = \mathbf{HY}$
- Matrix $\mathbf{H}$ used in diagnostics (Chapter 9)

Residuals

- Residual matrix

$$\begin{aligned}\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\ &= \mathbf{Y} - \mathbf{HY} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{Y}\end{aligned}$$

- $\mathbf{e}$ is a random vector

$$\begin{aligned}E(\mathbf{e}) &= (\mathbf{I} - \mathbf{H}) \times E(\mathbf{Y}) \\ &= (\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\ &= 0\end{aligned}$$

$$\begin{aligned}\boldsymbol{\sigma}^2(\mathbf{e}) &= (\mathbf{I} - \mathbf{H}) \times \boldsymbol{\sigma}^2(\mathbf{Y}) \times (\mathbf{I} - \mathbf{H})' \\ &= (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})' \\ &= (\mathbf{I} - \mathbf{H})\sigma^2\end{aligned}$$

ANOVA

- Quadratic form defined as

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_i \sum_j a_{ij} Y_i Y_j$$

where $\mathbf{A}$ is symmetric $n \times n$ matrix

- Sums of squares can be shown to be quadratic forms (page 207)

- Quadratic forms play a significant role in the theory of linear models when errors are normally distributed

Inference

- Vector $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{A}\mathbf{Y}$
- The mean and variance are

$$\begin{aligned} E(\mathbf{b}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\sigma}^2(\mathbf{b}) &= \mathbf{A} \times \boldsymbol{\sigma}^2(\mathbf{Y}) \times \mathbf{A}' \\ &= \mathbf{A} \times \sigma^2\mathbf{I} \times \mathbf{A}' \\ &= \sigma^2\mathbf{A}\mathbf{A}' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- Thus, $\mathbf{b}$ is *multivariate* Normal($\boldsymbol{\beta}$, $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$)

- Consider $\mathbf{X}_h' = [1 \quad X_h]$
- Mean response $\widehat{Y}_h = \mathbf{X}_h' b$

$$E(\widehat{Y}_h) = \mathbf{X}_h' \boldsymbol{\beta}$$

$$\text{Var}(\widehat{Y}_h) = \mathbf{X}_h' \times \boldsymbol{\sigma}^2(\mathbf{b}) \times \mathbf{X}_h = \sigma^2 \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h$$

- Prediction of new observation

$$\sigma^2\{pred\} = \sigma^2(1 + \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h)$$

$$s^2\{pred\} = MSE(1 + \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h)$$

Chapter Review

- Review of Matrices
- Regression Model in Matrix Form
- Calculations Using Matrices