Topic 7 - Matrix Approach to Simple Linear Regression

STAT 525 - Fall 2013

Outline

- Review of Matrices
- Regression model in matrix form
- Calculations using matrices

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

Matrix

- 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 is c=1

Matrix Operations

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

    Row 1 becomes Col 1, Row r becomes Col r

    $\downarrow$

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

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

Matrix Operations

- Multiplication
  - If scalar then $\lambda\mathbf{A} = [\lambda\mathbf{a_{ij}}]$
  - If multiplying two matrices ($\mathbf{AB}$)

    Cols of $\mathbf{A}$ must equal Rows of $\mathbf{B}$

    Resulting matrix of dimension Rows(A) $\times$ Col(B)
  - Elements obtained by taking cross products of rows of $\mathbf{A}$ with cols 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 $n = 4$

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

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

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

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

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 $\hat{\boldsymbol{\beta}}$ 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

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

Linear Dependence

- If there *is* a relationship between the column(row) vectors of a matrix such that $\lambda_1\mathbf{C}_1 + \ldots + \lambda_c\mathbf{C}_c = \mathbf{0}$ and not all λ's are 0, then the set of column(row) vectors are *linearly dependent*.
- If such a relationship does not exist then the set of column(row) vectors are *linearly independent.*
- Consider the matrix $\mathbf{Q}$ with column vectors $\mathbf{C}_1 - \mathbf{C}_3$

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

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

Rank of a Matrix

- Using $\lambda_1 = -2$, $\lambda_2 = 0$, $\lambda_3 = 1$

$$-2\begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 10 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- The columns of $\mathbf{Q}$ are *linearly dependent*
- The **rank** of a matrix is number of linear independent columns (or rows)
- Rank of a matrix cannot exceed $\min(r, c)$
- **Full Rank** $\equiv$ all columns are linearly independent
- In this example: The rank of $\mathbf{Q}$ is 2

Inverse of a matrix

- Inverse similar to the reciprocal of a scalar
- Inverse defined for square matrix of rank r
- Want to find the inverse of $\mathbf{S}$, 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

Inverse of a matrix

- 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 $\frac{1}{D}$

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

  - 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

$$\underset{(r \times r)}{\mathbf{A}} \quad \underset{(r \times 1)}{\mathbf{X}} = \underset{(r \times 1)}{\mathbf{C}}$$

- Assuming A has an inverse:

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

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: $E(\mathbf{A}) = [\mathbf{E}(\mathbf{A_i})]$
- Covariance matrix: symmetric

$$\boldsymbol{\sigma}^2(\mathbf{A}) = \begin{bmatrix} \sigma^2(A_1) & \sigma(A_1, A_2) & \cdots & \sigma(A_1, A_r) \\ \sigma(A_2, A_1) & \sigma^2(A_2) & \cdots & \sigma(A_2, A_r) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma(A_r, A_1) & \sigma(A_r, A_2) & \cdots & \sigma^2(A_r) \end{bmatrix}$$

Basic Theorems

- Consider random vector $\mathbf{Y}$
- Consider constant matrix $\mathbf{A}$
- Suppose $\mathbf{W} = \mathbf{AY}$
  - $\mathbf{W}$ is also a random vector
  - $E(\mathbf{W}) = \mathbf{AE}(\mathbf{Y})$
  - $\boldsymbol{\sigma}^2(\mathbf{W}) = \mathbf{A}\boldsymbol{\sigma^2}(\mathbf{Y})\mathbf{A}'$

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

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

- Taking derivative $\longrightarrow -2\mathbf{X}'\mathbf{Y} + \mathbf{2X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$
- This means $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$

Fitted Values

- The fitted values $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$
- Matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$ called *hat matrix*
- Equivalently write $\hat{\mathbf{Y}} = \mathbf{HY}$
- $\mathbf{H}$ symmetric and idempotent ($\mathbf{HH} = \mathbf{H}$)
- Matrix $\mathbf{H}$ used in diagnostics (chapter 9)

Residuals

- Residual matrix

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

- $\mathbf{e}$ a random vector

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

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

ANOVA

- Quadratic form defined as

$$\mathbf{Y'AY} = \sum_{\mathbf{i}} \sum_{\mathbf{j}} \mathbf{a_{ij} Y_i Y_j}$$

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

- Sums of squares can be shown to be quadratic forms (page 207)
- Quadratic forms play significant role in the theory of linear models when errors are Normally distributed

Inference

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

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

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

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

Inference Continued

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

  $E(\hat{Y}_h) = \mathbf{X_h'}\boldsymbol{\beta}$

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

- Prediction

  $E(\hat{Y}_h) = \mathbf{X_h'}\boldsymbol{\beta}$

  $\text{Var}(\hat{Y}_h) = \sigma^2(1 + \mathbf{X_h'}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X_h})$

Background Reading

- KNNL Chapter 5
- KNNL Sections 6.1-6.5