Matrix Approach to Linear Regresssion

Frank Wood

November 3, 2010

Random Vectors and Matrices

Let's say we have a vector consisting of three random variables

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

The expectation of a random vector is defined as

$$\mathbb{E}(\mathbf{y}) = \begin{pmatrix} \mathbb{E}(Y_1) \\ \mathbb{E}(Y_2) \\ \mathbb{E}(Y_3) \end{pmatrix}$$

Expectation of a Random Matrix

The expectation of a random matrix is defined similarly

$$\mathbb{E}(\mathbf{y}) = [\mathbb{E}(Y_{ij})] \qquad i = 1, \ldots n; j = 1, \ldots, p$$

Covariance Matrix of a Random Vector

The correlation of variances and covariances of and between the elements of a random vector can be collection into a matrix called the covariance matrix

$$cov(\mathbf{y}) = \sigma^2\{\mathbf{y}\} = \begin{pmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \sigma(Y_1, Y_3) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \sigma(Y_2, Y_3) \\ \sigma(Y_3, Y_1) & \sigma(Y_3, Y_2) & \sigma^2(Y_3) \end{pmatrix}$$

remember $\sigma(Y_2, Y_1) = \sigma(Y_1, Y_2)$ so the covariance matrix is symmetric

Derivation of Covariance Matrix

In vector terms the covariance matrix is defined by

$$\sigma^2\{\mathbf{y}\} = \mathbb{E}(\mathbf{y} - \mathbb{E}(\mathbf{y}))(\mathbf{y} - \mathbb{E}(\mathbf{y}))'$$

because

$$\sigma^2\{\mathbf{y}\} = \mathbb{E}(\begin{pmatrix} Y_1 - \mathbb{E}(Y_1) \\ Y_2 - \mathbb{E}(Y_2) \\ Y_3 - \mathbb{E}(Y_3) \end{pmatrix} \begin{pmatrix} Y_1 - \mathbb{E}(Y_1) & Y_2 - \mathbb{E}(Y_2) & Y_3 - \mathbb{E}(Y_3) \end{pmatrix})$$

Regression Example

- Take a regression example with $n = 3$ with constant error terms $\sigma^2\{\epsilon_i\}$ and are uncorrelated so that $\sigma^2\{\epsilon_i, \epsilon_j\} = 0$ for all $i \neq j$
- The covariance matrix for the random vector ϵ is

$$\boldsymbol{\sigma^2}\{\epsilon\} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

which can be written as $\boldsymbol{\sigma^2}\{\epsilon\} = \sigma^{\mathbf{2}}\ \mathbf{I}$

Basic Results

If $\mathbf{A}$ is a constant matrix and $\mathbf{y}$ is a random vector then $\mathbf{W} = \mathbf{A}\mathbf{y}$ is a random vector

$$\mathbb{E}(\mathbf{A}) = \mathbf{A}$$
$$\mathbb{E}(\mathbf{W}) = \mathbb{E}(\mathbf{A}\mathbf{y}) = \mathbf{A}\,\mathbb{E}(\mathbf{y})$$
$$\sigma^2\{\mathbf{W}\} = \sigma^2\{\mathbf{A}\mathbf{y}\} = \mathbf{A}\sigma^2\{\mathbf{y}\}\mathbf{A}'$$

Multivariate Normal Density

- Let Y be a vector of p observations

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_p \end{pmatrix}$$

- Let μ be a vector of p means of each of the p observations

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_p \end{pmatrix}$$

Multivariate Normal Density

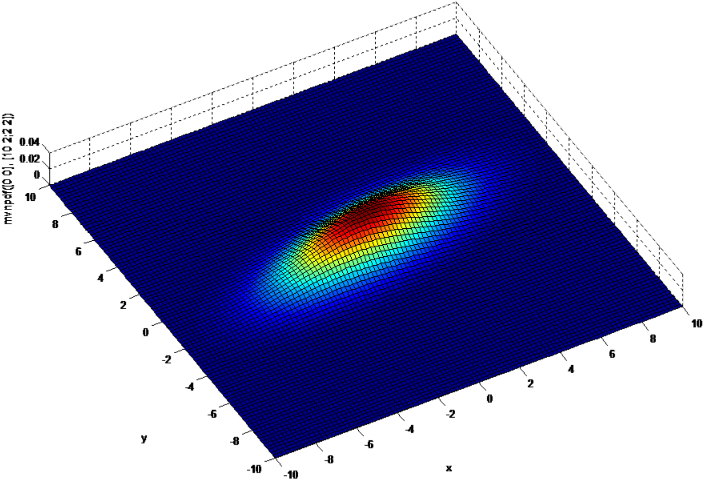
let Σ be the covariance matrix of Y

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ . & & & \\ . & & & \\ . & & & \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix}$$

Then the multivariate normal density is given by

$$P(\mathbf{Y}|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} exp(-\frac{1}{2}(\mathbf{Y} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mu))$$

Example 2d Multivariate Normal Distribution



Matrix Simple Linear Regression

- Nothing new-only matrix formalism for previous results
- Remember the normal error regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, ..., n$$

- Expanded out this looks like

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\ &\ldots \\ Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n \end{aligned}$$

- which points towards an obvious matrix formulation.

Regression Matrices

- If we identify the following matrices

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \\ \cdot & \\ \cdot & \\ 1 & X_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

- We can write the linear regression equations in a compact form $\mathbf{y} = \mathbf{X}\beta + \epsilon$

Regression Matrices

- Of course, in the normal regression model the expected value of each of the ϵ's is zero, we can write $\mathbb{E}(\mathbf{y}) = \mathbf{X}\beta$
- This is because

$$\mathbb{E}(\epsilon) = \mathbf{0}$$

$$\begin{pmatrix} \mathbb{E}(\epsilon_1) \\ \mathbb{E}(\epsilon_2) \\ . \\ . \\ . \\ \mathbb{E}(\epsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

Error Covariance

Because the error terms are independent and have constant variance σ^2

$$\sigma^2\{\epsilon\} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

$$\sigma^2\{\epsilon\} = \sigma^2\mathbf{I}$$

Matrix Normal Regression Model

In matrix terms the normal regression model can be written as

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

where $\mathbb{E}(\epsilon) = \mathbf{0}$ and $\sigma^2\{\epsilon\} = \sigma^2\mathbf{I}$, i.e. $\epsilon \sim N(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$

Least Square Estimation

If we remember both the starting normal equations that we derived

$$nb_0 + b_1 \sum X_i = \sum Y_i$$
$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i$$

and the fact that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_1 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & \\ . & \\ . & \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_1 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Least Square Estimation

Then we can see that these equations are equivalent to the following matrix operations

$$\mathbf{X}'\mathbf{X}\ \mathbf{b} = \mathbf{X}'\mathbf{y}$$

with

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$$

with the solution to this equation given by

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

when $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

When does $(\mathbf{X}'\mathbf{X})^{-1}$ exist?

$\mathbf{X}$ is an $n \times p$ (or $p+1$ depending on how you define p) design matrix.

$\mathbf{X}$ must have full column rank in order for the inverse to exist, i.e. $rank(\mathbf{X}) = p \implies (\mathbf{X}'\mathbf{X})^{-1}$ exists.