Introduction to Matrix Algebra

J. Alexander Branham

Fall 2015

Scalars

- Let's start with something familiar, with a new word

Scalars

- Let's start with something familiar, with a new word
- One number (12, for example) is referred to as a *scalar*

Scalars

- Let's start with something familiar, with a new word
- One number (12, for example) is referred to as a *scalar*
- This can be thought of as a 1x1 matrix

Scalars

- Let's start with something familiar, with a new word
- One number (12, for example) is referred to as a *scalar*
- This can be thought of as a 1*x*1 matrix
  - More on that in a bit...

$$\begin{bmatrix} 12 \end{bmatrix} = c$$

Vectors

- We can put several scalars together to make a vector

Vectors

- We can put several scalars together to make a vector
- For example:

$$\begin{bmatrix} 12 \\ 14 \\ 15 \end{bmatrix} = b$$

Vectors

- We can put several scalars together to make a vector
- For example:

$$\begin{bmatrix} 12 \\ 14 \\ 15 \end{bmatrix} = b$$

- Since this is a column of numbers, we cleverly refer to it as a *column vector*

Row Vectors

If we take b and arrange it so that it it a row of numbers instead of a column, we refer to it as a *row vector*:

$$\begin{bmatrix} 12 & 14 & 15 \end{bmatrix} = d$$

Matrix

We can put multiple vectors together to get a *matrix*:

$$\begin{bmatrix} 12 & 14 & 15 \\ 115 & 22 & 127 \\ 193 & 29 & 219 \end{bmatrix} = A$$

Matrices, cntd

- We refer to the *dimensions* of matrices by *row* x *column*

Matrices, cntd

- We refer to the *dimensions* of matrices by *row* x *column*
- So A is a $3x3$ matrix.

Matrices, cntd

- We refer to the *dimensions* of matrices by *row* x *column*
- So A is a 3x3 matrix.
- Note that matrices are usually designated by capital letters

Matrices, cntd

- We refer to the *dimensions* of matrices by *row* x *column*
- So A is a 3x3 matrix.
- Note that matrices are usually designated by capital letters
    - And sometimes bolded as well

Dimensions

ROW x COLUMN

Indices

- How do we refer to specific elements of the matrix???

Indices

- How do we refer to specific elements of the matrix???
- Solution: come up with a clever indexing scheme

Indices

- How do we refer to specific elements of the matrix???
- Solution: come up with a clever indexing scheme
- Matrix A is an mxn matrix where $m = n = 3$.

Indices

- How do we refer to specific elements of the matrix???
- Solution: come up with a clever indexing scheme
- Matrix A is an mxn matrix where $m = n = 3$.
- More generally, matrix B is an mxn matrix where the elements look like this:

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1m} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \dots & b_{nm} \end{bmatrix}$$

Addition and subtraction are EASY!

- Requirement: Must have *exactly* the same dimensions

Addition and subtraction are EASY!

- Requirement: Must have *exactly* the same dimensions
- To do the operation, just add or subtract each element with the corresponding element from the other matrix:

$$A \pm B$$

Addition and Subtraction

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & a_{13} \pm b_{13} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & a_{23} \pm b_{23} \\ a_{31} \pm b_{31} & a_{32} \pm b_{32} & a_{33} \pm b_{33} \end{bmatrix}$$

Scalar multiplication

Easy - just multiply each element of the matrix by the scalar

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ ca_{31} & ca_{32} & ca_{33} \end{bmatrix}$$

Matrix multiplication

- Requirement: the two matrices must be *conformable*

Matrix multiplication

- Requirement: the two matrices must be *conformable*
- This means that the number of columns in the first matrix equals the number of rows in the second

Matrix multiplication

- Requirement: the two matrices must be *conformable*
- This means that the number of columns in the first matrix equals the number of rows in the second
- The resulting matrix will have the number of rows in the first, and the number of columns in the second!

Pop quiz

- Which can we multiply? What will the resulting dimensions be?

$$b = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} M = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 2 & 4 \\ 2 & 3 & 2 \end{bmatrix} L = \begin{bmatrix} 6 & 5 & -1 \\ 1 & 4 & 3 \end{bmatrix}$$

Pop quiz

- Which can we multiply? What will the resulting dimensions be?

$$b = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} M = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 2 & 4 \\ 2 & 3 & 2 \end{bmatrix} L = \begin{bmatrix} 6 & 5 & -1 \\ 1 & 4 & 3 \end{bmatrix}$$

- **ONLY** LM and **NOT** ML

Pop quiz

- Which can we multiply? What will the resulting dimensions be?

$$b = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} M = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 2 & 4 \\ 2 & 3 & 2 \end{bmatrix} L = \begin{bmatrix} 6 & 5 & -1 \\ 1 & 4 & 3 \end{bmatrix}$$

- **ONLY** LM and **NOT** ML
- The dimensions will be 2x3

How to actually do this?

- Multiply each row by each column

How to actually do this?

- Multiply each row by each column
- (Board examples)

How to actually do this?

- Multiply each row by each column
- (Board examples)

How to actually do this?

- Multiply each row by each column
- (Board examples)

```
L <- matrix(c(6,5,-1, 1,4,3),
            nrow=2, byrow=TRUE)
M <- matrix(c(1,0,2, 1,2,4, 2,3,2),
            nrow=3, byrow=TRUE)
L%*%M
```

```
##      [,1] [,2] [,3]
## [1,]    9    7   30
## [2,]   11   17   24
```

Matrix Division

Matrix Division

HAHAHA... NOPE

Properties of operators

- Addition and subtraction

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication
  - $AB \neq BA$

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication
  - $AB \neq BA$
  - $A(BC) = (AB)C$

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication
  - $AB \neq BA$
  - $A(BC) = (AB)C$
  - $A(B + C) = AB + AC$

Properties of operators

- Addition and subtraction
  - Associative $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication
  - $AB \neq BA$
  - $A(BC) = (AB)C$
  - $A(B + C) = AB + AC$
  - $(A + B)C = AC + BC$

Qu'est-ce que c'est?

- Switch the rows and columns

```r
L
```

```
##      [,1] [,2] [,3]
## [1,]    6    5   -1
## [2,]    1    4    3
```

```r
t(L)
```

```
##      [,1] [,2]
## [1,]    6    1
## [2,]    5    4
```

Qu'est-ce que c'est?

- Switch the rows and columns
- So a *nxm* matrix becomes *mxn*

```
L
```

```
##       [,1] [,2] [,3]
## [1,]    6    5   -1
## [2,]    1    4    3
```

```
t(L)
```

```
##       [,1] [,2]
## [1,]    6    1
## [2,]    5    4
```

Qu'est-ce que c'est?

- Switch the rows and columns
- So a *nxm* matrix becomes *mxn*
- Typically denoted L' or L^T

```
L
```

```
##      [,1] [,2] [,3]
## [1,]    6    5   -1
## [2,]    1    4    3
```

```
t(L)
```

```
##      [,1] [,2]
## [1,]    6    1
## [2,]    5    4
```

Properties of transposing

- Matrix is *always* conformable for multiplication with its transpose in both directions

Properties of transposing

- Matrix is *always* conformable for multiplication with its transpose in both directions
- $(A \pm B)' = A' \pm B'$

Properties of transposing

- Matrix is *always* conformable for multiplication with its transpose in both directions
- $(A \pm B)' = A' \pm B'$
- $A'' = A$

Properties of transposing

- Matrix is *always* conformable for multiplication with its transpose in both directions
- $(A \pm B)' = A' \pm B'$
- $A'' = A$
- $(AB)' = B'A'$

Properties of transposing

- Matrix is *always* conformable for multiplication with its transpose in both directions
- $(A \pm B)' = A' \pm B'$
- $A'' = A$
- $(AB)' = B'A'$
- $(cA)' = cA'$ where c is a scalar

Matrix Inverses

- I kinda lied when I said there isn't matrix division

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an nxn square matrix:

$$AB = BA = I_n$$

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an *nxn* square matrix:

$$AB = BA = I_n$$

- Then B is said to be the *inverse* of A

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an *nxn* square matrix:

$$AB = BA = I_n$$

- Then B is said to be the *inverse* of A
  - This is usually denoted A^{-1}

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an *nxn* square matrix:

$$AB = BA = I_n$$

- Then B is said to be the *inverse* of A
  - This is usually denoted A^{-1}
  - So $AA^{-1} = I_n$

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an *nxn* square matrix:

$$AB = BA = I_n$$

- Then B is said to be the *inverse* of A
  - This is usually denoted A^{-1}
  - So $AA^{-1} = I_n$
- If B doesn't exist, then the matrix is *singluar*

Matrix Inverses

- I kinda lied when I said there isn't matrix division
- We use matrix inverses all the time
- If A is an *nxn* square matrix:

$$AB = BA = I_n$$

- Then B is said to be the *inverse* of A
    - This is usually denoted A^{-1}
    - So $AA^{-1} = I_n$
- If B doesn't exist, then the matrix is *singluar*
- Finding inverses by hand is super hard (especially as n increases), so we let computers do this for us

Some properties

- Let A be nxn square matrix:

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:
- A is full rank: $rank(A) = n$

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:
- A is full rank: $rank(A) = n$
- A' is also invertible

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:
- A is full rank: $rank(A) = n$
- A' is also invertible
- $(A^{-1})^{-1} = A$

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:
- A is full rank: $rank(A) = n$
- A' is also invertible
- $(A^{-1})^{-1} = A$
- $(cA)^{-1} = c^{-1}A^{-1}$ for nonzero scalar k

Some properties

- Let A be nxn square matrix:
- If A^{-1} exists:
- A is full rank: $rank(A) = n$
- A' is also invertible
- $(A^{-1})^{-1} = A$
- $(cA)^{-1} = c^{-1}A^{-1}$ for nonzero scalar k
- $(A')^{-1} = (A^{-1})'$

Special types of matrices

Some matricies get more love than others

Square matrix

Any *nxn* matrix (same number rows and columns)

$$\begin{bmatrix} 1 & 0 & 2 \\ 1 & 2 & 4 \\ 2 & 3 & 2 \end{bmatrix}$$

Symmetric matrix

A square matrix that is the same as its transpose

$$\begin{bmatrix} 2 & 5 & 7 \\ 5 & 9 & 6 \\ 7 & 6 & 7 \end{bmatrix}$$

Diagonal matrix

A symmetric matrix with zeros everywhere but the main diagonal

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

Scalar matrix

A diagonal matrix with the same number all along the diagonal

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Identity matrix

- A scalar matrix where the diagonal elements are 1.

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

Identity matrix

- A scalar matrix where the diagonal elements are 1.

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

Identity matrix

- A scalar matrix where the diagonal elements are 1.

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

- This is a super important type of matrix.

Identity matrix

- A scalar matrix where the diagonal elements are 1.

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

- This is a super important type of matrix.
- It gets its own notation: I_n where n is the number of rows and columns

Identity matrix

- A scalar matrix where the diagonal elements are 1.

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

- This is a super important type of matrix.
- It gets its own notation: I_n where n is the number of rows and columns
- Note that $I_nA = A$ and also $AI_n = A$

Notation

- Let's pretend that we know the **true** model

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix
  - β is $kx1$ column vector

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix
  - β is $kx1$ column vector
  - E is $nx1$ column vector

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix
  - β is $kx1$ column vector
  - E is $nx1$ column vector
- Therefore, we have:

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix
  - β is $kx1$ column vector
  - E is $nx1$ column vector
- Therefore, we have:

Notation

- Let's pretend that we know the **true** model
  - Y is $nx1$ column vector
  - X is $nxk(+1)$ matrix
  - β is $kx1$ column vector
  - E is $nx1$ column vector
- Therefore, we have:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ 1 & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Matrix form

$$Y = X\beta + E$$

OLS minimizes the sum of squared residuals

- How to do that in matrix form?

OLS **minimizes the sum of squared residuals**

- How to do that in matrix form?
- First, what is sum of squared residuals?

OLS **minimizes the sum of squared residuals**

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

OLS **minimizes the sum of squared residuals**

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

OLS minimizes the sum of squared residuals

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

- (show why on board)

OLS minimizes the sum of squared residuals

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

- (show why on board)
- Alternatively,

$$\begin{aligned} E'E &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - \hat{\beta}X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta} \end{aligned}$$

To **minimize** the sum of squares, we take the derivative

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta}$$

To **minimize** the sum of squares, we take the derivative

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

To **minimize** the sum of squares, we take the derivative

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

- To check that this is a minimum, we check to make sure that the second derivative is positive

To **minimize** the sum of squares, we take the derivative

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

- To check that this is a minimum, we check to make sure that the second derivative is positive
- The second derivative is $2X'X$, which is positive definite so long as X is full rank

Solve for the estimator

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

Solve for the estimator

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

Solve for the estimator

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

Solve for the estimator

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

- We know that $(X'X)^{-1}(X'X) = I$

$$(X'X)^{-1}X'Y = I\hat{\beta}$$

Solve for the estimator

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

- We know that $(X'X)^{-1}(X'X) = I$

$$(X'X)^{-1}X'Y = I\hat{\beta}$$

- And I is (kinda) like multiplying by 1 so :

$$(X'X)^{-1}X'Y = \hat{\beta}$$