Political Science 236

Review of Matrix Algebra

Rocio Titiunik

Fall 2007

1 Matrix Algebra

In this section, I will cover the basics of matrix algebra. Let's first start by defining what a matrix is.

Definition 1.1 *Matrix. A matrix is a rectangular array of elements. The size or dimension of a matrix is given by the number of its rows and the number of its columns.*

There is some standard notation to deal with matrices. A matrix with k rows and n columns is called a $k \times n$ ("k by n") matrix, and the element in row i and column j is written a_{ij} and referred to as "the $(i, j)\,th$ entry or element".

An abstract representation of a $k \times n$ matrix A is:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{pmatrix}$$

Provided the dimensions are what they need to be, two matrices can be added, subtracted, multiplied and divided. I will describe these and other operations in the next couple of pages.

First, keep in mind that two matrices A and B are equal if they both have the same dimension and if the corresponding entries are equal (i.e. if $a_{ij} = b_{ij}$ for all i, j).

1.1 Addition

Matrices can be added if they have the same dimensions, i.e. the same number of rows and columns. Their sum is a matrix with the same dimensions than the two matrices being added, and the $(i, j)\,th$ entry of the sum matrix is simply the sum of the two $(i, j)\,th$ entries of the two matrices being added. This, of course, can be generalized to the addition of any number of matrices.

Symbolically,

$$\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & ... & b_{1n} \\ b_{21} & b_{22} & ... & b_{2n} \\ ... & ... & ... & ... \\ b_{k1} & b_{k2} & ... & b_{kn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & ... & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & ... & a_{2n}+b_{2n} \\ ... & ... & ... & ... \\ a_{k1}+b_{k1} & a_{k2}+b_{k2} & ... & a_{kn}+b_{kn} \end{pmatrix}$$

An example would be:

$$\begin{pmatrix} 3 & 4 & 9 \\ 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 4 & 7 & 9 \\ 0 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 7 & 11 & 18 \\ 0 & 4 & 2 \end{pmatrix}$$

However, the following sum

$$\begin{pmatrix} 3 & 4 & 9 \\ 0 & 1 & 1 \\ 1 & 4 & 6 \end{pmatrix} + \begin{pmatrix} 4 & 7 & 9 \\ 0 & 3 & 1 \end{pmatrix}$$

is not defined.

If we define by $\mathbf{0}$ the matrix whose elements are all zero, we have $A + \mathbf{0} = A$ for any matrix A.

This is, the matrix **0** is the *additive identity*. Symbolically,

$$\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} + \begin{pmatrix} 0_{11} & 0_{12} & ... & 0_{1n} \\ 0_{21} & 0_{22} & ... & 0_{2n} \\ ... & ... & ... & ... \\ 0_{k1} & 0_{k2} & ... & 0_{kn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix}$$

1.2 Subtraction

Once we know how to add matrices, the only thing that we need to know in order to subtract is how to define the opposite of a matrix, i.e. "$-A$". Since $-A$ is what we add to A in order to obtain **0**, we have:

$$-\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} = \begin{pmatrix} -a_{11} & -a_{12} & ... & -a_{1n} \\ -a_{21} & -a_{22} & ... & -a_{2n} \\ ... & ... & ... & ... \\ -a_{k1} & -a_{k2} & ... & -a_{kn} \end{pmatrix}$$

So $-A$ is just the matrix A where all the elements have the opposite sign than they had in A.

Since $A - B$ is actually $A + (-B)$, subtraction is just

$$\begin{aligned} A - B &= A + (-B) \\ &= \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} + \begin{pmatrix} -b_{11} & -b_{12} & ... & -b_{1n} \\ -b_{21} & -b_{22} & ... & -b_{2n} \\ ... & ... & ... & ... \\ -b_{k1} & -b_{k2} & ... & -b_{kn} \end{pmatrix} = \\ &= \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} & ... & a_{1n} - b_{1n} \\ a_{21} - b_{21} & a_{22} - b_{22} & ... & a_{2n} - b_{2n} \\ ... & ... & ... & ... \\ a_{k1} - b_{k1} & a_{k2} - b_{k2} & ... & a_{kn} - b_{kn} \end{pmatrix} \end{aligned}$$

1.3 Scalar Multiplication

Matrix can be multiplied by scalars, i.e. ordinary numbers. The operation of multiplying a matrix by a number is called scalar multiplication. The product of a matrix A and the number r is denoted by rA and is the matrix created by multiplying each element of A by the number r. Symbolically,

$$rA = r\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} = \begin{pmatrix} ra_{11} & ra_{12} & ... & ra_{1n} \\ ra_{21} & ra_{22} & ... & ra_{2n} \\ ... & ... & ... & ... \\ ra_{k1} & ra_{k2} & ... & ra_{kn} \end{pmatrix}$$

Note that we implicitly used this definition of scalar multiplication above when we defined $-A$ as $(-1) \times A$.

1.4 Matrix Multiplication

Matrices can be multiplied together, but only if they have the right dimensions. Two matrices A and B can multiplied if and only if:

$$\text{Number of columns of } A = \text{Number of rows of } B$$

This means that for the product AB to exist A must be $k \times m$ and B must be $m \times n$. To obtain $(i, j)\,th$ entry of the product matrix AB we must multiply the ith row of A and the jth column of B as follows:

$$(a_{i1}, a_{i2}, ..., a_{im}) \cdot \begin{pmatrix} b_{1j} \\ b_{2j} \\ ... \\ b_{mj} \end{pmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{im}b_{mj} =$$

which can be also written as

$$(i.j)\,th \text{ entry of } AB = \sum_{h=1}^{m} a_{ih}b_{hj}$$

For example,

$$\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 3\cdot 1+2\cdot 4 & 3\cdot 2+2\cdot 5 & 3\cdot 3+2\cdot 6 \\ 1\cdot 1+4\cdot 4 & 1\cdot 2+4\cdot 5 & 1\cdot 3+4\cdot 6 \end{pmatrix}$$
$$= \begin{pmatrix} 11 & 16 & 21 \\ 17 & 22 & 27 \end{pmatrix}$$

Note that in this case

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$$

is not defined.

In general, if A is $(k \times m)$ and B is $(m \times n)$, the product AB will be $(k \times n)$. In other words,

$$\begin{aligned} \text{Number of rows of } AB &= \text{Number of columns of } A \\ \text{Number of columns of } AB &= \text{Number of columns of } B \\ (k \times m) \cdot (m \times n) &= (k \times n) \end{aligned}$$

Note that, in general, we'll have:

$$AB \neq BA$$

This is trivially true when one product is defined and the other is not. But we can have $AB \neq BA$ *even when both products are defined.* This means that the product of matrices is *not commutative.* The following is an example of this non-commutative property of matrix multiplication.

Example 1.1 *Matrices do not satisfy the commutative law of multiplication.*

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

while

$$\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}$$

Before we leave this section, let me define the $(n \times n)$ matrix I. This matrix is defined as

$$I=\begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix}$$

with elements $a_{ii} = 1$ for all i and $a_{ij} = 0$ for all $i \neq j$. This matrix has the property that for any $(m \times n)$ matrix A,

$$AI=A$$

and for any $(n \times x)$ matrix B,

$$IB=B$$

The matrix I is called the *n-identity matrix* because it is a multiplicative identity for matrices in the same way the number 1 is for real numbers.

1.5 Laws of Matrix Algebra

1. Associative Laws:

$$(A+B)+C = A+(B+C)$$

$$(AB)\,C = A\,(BC)$$

2. Commutative Law for Addition:

$$A+B = B+A$$

3. Distributive Laws:

$$A\,(B+C) = AB+AC$$

$$(A+B)\,C = AC+BC$$

4. Commutative Law for Multiplication: DOES NOT HOLD!!!!

1.6 Transpose

There is one more matrix operation that is used very often, the transpose. The transpose of $(k \times n)$ matrix A is the $(n \times k)$ matrix obtained by interchanging the rows and columns of A. The transpose of A is denoted by A^T. The first row of A becomes the first columns of A^T, the second row of A becomes the second column of A^T, etc. In general, the $(i, j)\,th$ entry of A becomes the $(j, i)\,th$ entry of A^T. For example,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}$$

The following rules can be easily proved for any $(k \times n)$ matrices A and B and any scalar r:

$$\begin{aligned} (A+B)^T &= A^T + B^T \\ (A-B)^T &= A^T - B^T \\ \left(A^T\right)^T &= A \\ (rA)^T &= rA^T \\ (AB)^T &= B^T A^T \end{aligned}$$

1.7 Rank

Before introducing the concept of rank let me introduce the concept of linear dependence of vectors.

Definition 1.2 *Linear Dependence. Vectors* $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ *are linearly dependent in* $\mathbb{R}^n$ *if and only if there exists numbers* $c_1, c_2, ..., c_n$ ***not all zero*** *such that*

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = \mathbf{0}_n$$

Whenever this numbers do not exist, we say that the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are linearly independent. In other words, the vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are linearly independent if the only numbers $c_1, c_2, ..., c_n$ that satisfy $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = 0$ are all zero, i.e. $c_1 = c_2 = ... = c_n = 0$.

Example 1.2 *The vectors* $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ *are linearly dependent because* $c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *holds for non-zero* c_1 *and* c_2. *To see why, notice that this can be expressed as* $\begin{pmatrix} c_1 \\ 2c_1 \end{pmatrix} + \begin{pmatrix} 2c_2 \\ 4c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. *This is the same as the system*

$$\begin{aligned} c_1 + 2c_2 &= 0 \\ 2c_1 + 4c_2 &= 0 \end{aligned}$$

but note that the second equation is just the first equation multiplied by two. So dividing both sides of the second equation by two we get $c_1 + 2c_2 = 0$ *and hence the system is reduced to one equation. This implies that* ***any*** c_1 *and* c_2 *such that* $c_1 = -2c_2$ *will make* $c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. *This means that the vectors* $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ *are linearly dependent.*

As this example shows, when one vector can be written as a linear combination of another vector, the vectors will always be linearly dependent.

Example 1.3 *Independent vectors. The vectors* $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ *and* $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ *are linearly independent To show this, it suffices to show that the system* $c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *has the unique solution* $c_1 = c_2 = 0$. *This is in fact the case, since this can be expressed as the system*

$$\begin{aligned} c_1 + -3c_2 &= 0 \\ c_1 + 2c_2 &= 0 \end{aligned}$$

which yields $c_1 = c_2 = 0$.

Now that we know what linear independence is, we can define the rank of a matrix.

Definition 1.3 *The rank of a square matrix A is the maximum number of columns (or rows) of A which are linearly independent.*

Definition 1.4 *Full-rank. If the rank of a square matrix A is equal the number of its columns (or rows), we say that A is **full-rank**.*

It can be shown that for any matrix A, $rank\left(A^T A\right) = rank\left(A\right)$. We will use this property in the matrix version of OLS.

2 Systems of Equations in Matrix Form

The algebraic operations for matrices described thus far can be used to express systems of linear equations in what we call "matrix form". Consider a typical system of linear equations where x_i are unknown variables, and a_{ij}'s and b_i's are real numbers:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 \\ &... \\ a_{k1}x_1 + a_{k2}x_2 + ... + a_{kn}x_n &= b_k \end{aligned}$$

A solution to this system is an n-tuple of real numbers $x_1, x_2, ..., x_n$ that satisfies each of these k equations. For a linear system such as this we are interested in (i) whether a solution exists, (ii) how many solutions there are, and (iii) an algorithm to compute the solutions. Matrix algebra is one the methods that can be used to solve these systems.

We can express this system using matrix algebra as follows. Let A denote the $(k \times n)$ coefficient matrix of the system:

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix}$$

and let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_k \end{pmatrix}$$

The system of equations can be written as

$$\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ ... \\ b_k \end{pmatrix}$$

or even more compactly as

$$A\mathbf{x} = \mathbf{b}$$

where $A\mathbf{x}$ refers to the product of the $(k \times n)$ matrix A with the $(n \times 1)$ matrix $\mathbf{x}$ This product is a $(k \times 1)$ matrix, which must be made equal to the $(k \times 1)$ matrix $\mathbf{b}$.

2.1 Special Matrices

Let A be an arbitrary $(k \times n)$ matrix.

1. **Square Matrix**. A is square if $k = n$
2. **Column Matrix**. A is a column matrix if $n = 1$
3. **Row Matrix**. A is a row matrix if $k = 1$

4. **Diagonal Matrix**. A is a diagonal matrix if $k = n$ and $a_{ij} = 0$ for all $i \neq j$ (a square matrix in which all non-diagonal elements are zero)

5. **Symmetric Matrix**. A is symmetric if $A^T = A$, i.e. $a_{ij} = a_{ji}$ for all i, j. Symmetric matrices must be square.

6. **Idempotent Matrix.** A is idempotent if $AA = A$. Idempotent matrices must be square (or else the product is not conformable).

7. **Nonsingular Matrix.** The coefficient matrix A is non-singular if the corresponding linear system $A\mathbf{x} = \mathbf{b}$ has one and only one solution.

8. **Full-rank Matrix**. The square matrix A is full rank if its rank is equal to the number of its columns (rows).

2.2 Using vectors to express sums

Let $\mathbf{x}$ be an $(n \times 1)$ column vector. If we denote by $\mathbf{i}$ the column vector $(n \times 1)$ whose elements are all ones ,then

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n = \mathbf{i}^T \mathbf{x}$$

and

$$\mathbf{i}^T \mathbf{i} = n$$

It follows that for any constant a we have

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i = a \mathbf{i}^T \mathbf{x}$$

and if $a = \frac{1}{n}$ we have

$$\overline{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \mathbf{i}^T \mathbf{x}$$

The sum of squares of the vector $\mathbf{x}$ is

$$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + ... + x_n^2 = \mathbf{x}^T\mathbf{x}$$

and the sum of the products of the n elements in vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + ... + x_n y_n = \mathbf{x}^T\mathbf{y}$$

3 Algebra of Square Matrices

All the operations defined above hold for square matrices. The sum, subtraction, and multiplication of $(n \times n)$ squares matrices is $(n \times n)$. What about dividing matrices? For scalars, dividing by x is just multiplying by $\frac{1}{x} = x^{-1}$, which is defined as long as $x \neq 0$. To "divide" matrices, we need to specify what the inverse of a matrix (written A^{-1}) really means. So let's see what it means.

Definition 3.1 *Inverse of a matrix. Let A be an $(n \times n)$ matrix. The $(n \times n)$ matrix B is an inverse for A if $AB = BA = I$.*

If this matrix B exists, we say that A is **invertible. (**Also, it is easy to show that any square matrix can have at most one inverse). If A is an $(n \times n)$ invertible matrix, we refer to its unique inverse as A^{-1}.

The following are very important theorems.

Theorem 3.1 *If an $(n \times n)$ matrix is invertible, then it is non-singular, and the unique solution to the system of equations $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$.*

Theorem 3.2 *An $(n \times n)$ matrix is non-singular if and only if it is invertible.*

We can summarize these relations between rank, invertibility, non-singularity and solution to the linear system in the following theorem.

Theorem 3.3 *For any square $(n \times n)$ matrix A, the following statements are equivalent:*

1. *A is invertible*

2. *Every system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$, which is $\mathbf{x} = A^{-1}\mathbf{b}$.*

3. *A is non-singular*

4. *A has full rank*

5. *The determinant of A is non-zero.*

The following theorem gives the properties of inverses.

Theorem 3.4 *Let A and B be square invertible matrices, Then*

- $\left(A^{-1}\right)^{-1} = A$
- $\left(A^{T}\right)^{-1} = \left(A^{-1}\right)^{T}$
- *AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$*

4 OLS in Matrix Form

The multiple linear regression model for unit i can be written as:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$$

and since we have n observations, we have the system:

$$\begin{aligned}
y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + \varepsilon_1 \\
y_i &= \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + \varepsilon_2 \\
&\quad ... \\
y_i &= \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \varepsilon_i
\end{aligned}$$

This system can be expressed in matrix form as

$$\begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1k} \\ x_{21} & x_{22} & ... & x_{2k} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ ... \\ \varepsilon_k \end{pmatrix}$$

where $\mathbf{y}$ is $(n \times 1)$, X is $(n \times k)$, β is $(k \times 1)$ and ε is $(n \times 1)$.

So the multiple regression model can be written as

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

In this notation, the sum of squared residuals is

$$\varepsilon^T \varepsilon = \sum_{i=1}^{n} \varepsilon_i^2$$

and the vector $\widehat{\beta}$ that minimizes $\varepsilon^T \varepsilon$ is

$$\begin{aligned} \widehat{\beta} &= \arg\min \left\{ \varepsilon^T \varepsilon \right\} \\ &= \arg\min \left\{ (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) \right\} \\ &= \arg\min \left\{ \mathbf{y}'\mathbf{y} - \mathbf{2}X^T \mathbf{y}\beta + \beta X^T X\beta \right\} \end{aligned}$$

The first order conditions are:

$$-2X^T \mathbf{y} + 2X^{\mathbf{T}} X\widehat{\beta} = \mathbf{0}$$

and if the matrix $X^T X$ is invertible, we have

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^T \mathbf{y}$$