Matrix Algebra

A matrix is a rectangular array (arrangement) of real numbers. The number of rows and columns may very from one matrix to another, so we conveniently describe the size of a matrix by giving its dimensions—that is, the number of its rows and columns. For example, matrix **A** consisting of two rows and three columns is written as

$$\mathbf{A}_{2\times 3} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}.$$

Denote a_{ij} and b_{ij} to be the scalar in the ith row and jth column of matrix **A** and **B**, respectively. Matrices **A** and **B** are equal if and only if they have the same dimensions and each element of **A** equals the corresponding element of **B**, i.e. $a_{ij} = b_{ij}$. The **transpose** of a matrix **A**, denoted $\mathbf{A}^{\mathrm{T}}$ (or $\mathbf{A}'$), is obtained by creating the matrix whose kth row is the kth column of the original matrix. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \qquad \mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

Suppose that **A** is an $n \times m$ matrix. If n equals m, then **A** is a **square** matrix. There are several particular types of square matrices:

(1) Matrix **A** is **symmetric** if $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$.

(2) A **diagonal matrix** is a square matrix whose only nonzero elements appear on the main diagonal, moving from upper left to lower right.

(3) A **scalar matrix** is a diagonal matrix with same value in all diagonal elements.

(4) An **identity matrix** is a scalar matrix with ones on the diagonal. This is always denoted **I**. A subscript is sometimes included to indicate its size, or **order**. For example,

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(5) A **triangular matrix** is one that has only zeros either above or below the main diagonal. If the zeros are above the diagonal, the matrix is lower triangular.

Two matrices, say **A** and **B**, can be added only if they are of the same dimensions. The sum of two matrices will be a matrix obtained by adding corresponding elements of matrices **A** and **B**—that is, elements in corresponding positions; that is, if $\mathbf{P} = \mathbf{A} + \mathbf{B}$, then $p_{ij} = a_{ij} + b_{ij}$ for all i and j.. For example,

$$\mathbf{A}_{2\times3} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{2\times3} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \end{bmatrix}$$

$$\mathbf{P} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 1+2 & 3+2 & 5+3 \\ 2+0 & 4+(-1) & 6+(-2) \end{bmatrix} = \begin{bmatrix} 3 & 5 & 8 \\ 2 & 3 & 4 \end{bmatrix}$$

The difference of matrices is carried out in a similar manner. For example, **Q** = **A** – **B**, then $q_{ij} = a_{ij} - b_{ij}$ for all i and j.

For an $m \times n$ matrix **A** and scalar c, we define the **product** of c and **A**, denoted c**A,** to be the $m \times n$ matrix whose ith row and jth column is $c\, a_{ij}$. We could denote (–1) **B** by – **B** and define the difference of **A** and **B**, denoted **A** – **B**, as **A** + (– **B**).

Let **A** be an $m \times \boldsymbol{n}$ matrix and let **B** be an $\boldsymbol{n} \times p$ matrix. We define the **matrix product** of **A** and **B** to be the $\boldsymbol{m} \times \boldsymbol{p}$ matrix **AB** whose ith row and jth column is the dot product of ith row of **A** and jth column of **B**. That is, the ith row and jth column of **AB** is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}\ .$$

Example: $\mathbf{A}_{2\times3} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$, and $\mathbf{B}_{3\times1} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. Then $\mathbf{AB}_{2\times1} = \begin{bmatrix} 1\times1+3\times0+5\times1=6 \\ 2\times1+4\times0+6\times1=8 \end{bmatrix}$.

Note that matrices can only be multiplied together if they are **conformable**. That is, we can only form the matrix product **AB** *if the number of column in* **A** *equals the number of rows in* **B**. Thus, although it is the case that in normal scalar algebra $ab = ba$, a similar property does not normally hold for matrices. That is, in general, **AB** ≠ **BA**.

The inverse of a square matrix **A** is defined as that matrix, written as $\mathbf{A}^{-1}$, for which

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}.$$

Note that only square matrices have inverses and that the inverse matrix is also square, otherwise it would be impossible to form both the matrix products $\mathbf{A}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-1}$. The inverse matrix is analogous to the reciprocal of ordinary scalar algebra. In order to find $\mathbf{A}^{-1}$, we first need to induce the concepts of determinants, minor, and cofactors.

All square matrices **A** have associated with them a scalar quantity (i.e. a number) known as the **determinant** of the matrix and denoted either by det (**A**) or by |**A**|. I believe that you can find the determinant of a 2 × 2 matrix without difficulties. Before we undertake the evaluation of larger determinants, it is necessary to define the term "**minor**" and "**cofactor**." If the ith row and jth column of a square matrix **A** are deleted, we obtain the so-called submatrix of the scalar a_{ij}. The determinant of this submatrix is known as its **minor**, and we denoted it by m_{ij}. All scalars in a matrix will obviously have minors.

The **cofactor** of any scalar in a matrix is closely related to the minor. The cofactor of the scalar a_{ij} in the matrix A is defined as

$$c_{ij} = (-1)^{i+j}\, m_{ij}$$

where m_{ij} is the corresponding minor. Thus, a cofactor is simply a minor with an appropriate sign attached to it.

The determinant of a square matrix with order 3 (or higher order) can be obtained by taking *any row* or *column* of scalars from the matrix, multiply each scalar by its respective cofactor, and sum the resultant products. For example, to find the determinant of the matrix **A** with order 3, we could expand using the first row. This could give

$$\det(\mathbf{A}) = a_{11}\, c_{11} + a_{12}\, c_{12} + a_{13}\, c_{13} = a_{11}\, m_{11} - a_{12}\, m_{12} + a_{13}\, m_{13}.$$

Notice that whether we expand by the first row or the second column, we obtain the same value for the determinant.

<u>Some properties of determinants</u>:

1. Let **A** be an upper (or lower) triangular matrix, then the determinant of **A** is the product of its diagonal entries.
2. If **A** is diagonal matrix, then its determinant is the product of its diagonal entries.
3. If two rows of a square matrix **A** are the same, det (**A**) = 0.
4. Let **A** and **B** be square matrices with the same order. Then det (**AB**) = det (**A**) det (**B**).
5. For any nonzero scalar a and a square matrix **A** with order n, $\det(a\mathbf{A}) = a^n \det(\mathbf{A})$.
6. Let **A** be a square matrix. Then **A** is invertible if and only if $\det(\mathbf{A}) \neq 0$.
7. If a square matrix **A** is invertible, then $\det(\mathbf{A}^{-1}) = 1 / \det(\mathbf{A})$.
8. $\det(\mathbf{A}^{T}) = \det(\mathbf{A})$.
9. Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be square matrices, we have $\det\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} = \det(\mathbf{A}_1)\det(\mathbf{A}_2)$.
10. $\det\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) = \det(\mathbf{A}_{11})\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})$.

Let **A** be the square matrix with order n. The **adjoint** of **A** is obtained by replacing each scalar a_{ij} by its cofactor c_{ij}, and then **<u>transposing</u>** it. That is,

$$adj(\mathbf{A}) = \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{11} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}.$$

Suppose that $\det(\mathbf{A}) \neq 0$. The inverse matrix $\mathbf{A}^{-1}$ can be shown to be

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} adj(\mathbf{A}) = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}.$$

Notice that the above equation is valid only if det ($\mathbf{A}$) ≠ 0. If det ($\mathbf{A}$) = 0, then the matrix **A** is said to be **singular** and *doe not have an inverse*. Furthermore, $\mathbf{A}^{-1}$**,** if it exists, is **unique**.

Example: $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix}$**.** Its determinant and adjoint matrix is, respectively,

$$\det(\mathbf{A}) = 6 \quad \text{and} \quad adj\,(\mathbf{A}) = \begin{bmatrix} -1 & 2 & 1 \\ 5 & -4 & 1 \\ 2 & 2 & -2 \end{bmatrix}^{\mathbf{T}} = \begin{bmatrix} -1 & 5 & 2 \\ 2 & -4 & 2 \\ 1 & 1 & -2 \end{bmatrix}.$$

<u>Some properties of the inverse</u>:

1. For any nonsingular matrices **A** and **B**, $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
2. If a square matrix **A** is invertible, then det ($\mathbf{A}^{-1}$) = 1 / det (**A**).
3. For any nonzero scalar a and nonsingular matrix **A**, $(a\,\mathbf{A})^{-1} = \frac{1}{a}\sum_{i=1}^{n} \mathbf{A}^{-1}$.
4. For any nonsingular matrix **A**, $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$.
5. For any nonsingular matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, $\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^{-1} \end{bmatrix}$
6. Given $\mathbf{Ax} = \boldsymbol{b}$, for any nonsingular matrix **A**, and two column vectors **x** and ***b***, then can solve $\mathbf{x} = \mathbf{A}^{-1}\boldsymbol{b}$.

The Trace of a Matrix

The trace tr(**A**) of a square matrix **A** is the sum of all diagonal elements.

1. tr($\mathbf{A}^{T}$) = tr(**A**).
2. tr(scalar) = tr(scalar).
3. Let **A** and **B** be square matrices with the same order. tr(**A + B**)= tr(**A**) **+** tr(**B**).
4. For any n×k matrix **A** and k×n matrix **B**, tr(**AB**)= tr(**BA**).
5. For any scalar a and square matrix **A**, tr(a**A**)= a·tr(**A**).
6. For conformable matrices **A**, **B, C,** and **D,** tr(**ABCD**) = tr(**DABC**) = tr(**CDAB**)

The Summation Vector

Denote $\mathbf{1}_n$ as an n-dimensional column vector with all elements 1, called the summation vector because it helps sum the elements of another vector $\mathbf{x}$ in a vector multiplication as follows:

$$\mathbf{1}_n'\mathbf{x} = [1 \quad 1 \quad \cdots \quad 1]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i \ .$$

Note that the result is a scalar, i.e., a single number.

The summation vector can be used to construct some interesting matrices as follows:

1. Since $\mathbf{1}_n'\mathbf{1}_n = n$ and $(\mathbf{1}_n'\mathbf{1}_n)^{-1} = 1/n$, we have
$$(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ ,$$
which is the average of the elements of the vector $\mathbf{x}$ and is usually denoted as $\bar{x}$.
2. We can "expand" the scalar $\bar{x}$ to a vector of $\bar{x}$ by simply multiplying it with $\mathbf{1}_n$ as follows:
$$\mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\mathbf{x} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\mathbf{1}_n = \bar{x}\cdot\mathbf{1}_n = \begin{bmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix}.$$
3. By subtracting the above vector of the average from the original vector $\mathbf{x}$, we get the following vector of deviations:
$$\mathbf{x} - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\mathbf{x} = \mathbf{x} - \bar{x}\cdot\mathbf{1}_n = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} = \left(\mathbf{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\right)\mathbf{x} = \mathbf{M}^0\mathbf{x}$$

where $\mathbf{M}^0 = \mathbf{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'$ is a square matrices with order n. The diagonal elements of $\mathbf{M}^0$ are all $(1-1/n)$ and its off-diagonal elements are all $-1/n$. Some useful results:

1. $\mathbf{M}^0\mathbf{1}_n = (\mathbf{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n')\,\mathbf{1}_n = \mathbf{1}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\,\mathbf{1}_n = \mathbf{0}$. Hence, $\mathbf{1}_n'\mathbf{M}^0 = \mathbf{0}'$. The sum of deviation about the mean is then $\sum_{i=1}^{n}(x_i - \bar{x}) = \mathbf{1}_n'(\mathbf{M}^0\mathbf{x}) = \mathbf{1}_n'\mathbf{0} = 0$.
2. The sum of squared deviations about the mean is
$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (\mathbf{x} - \bar{x}\mathbf{1}_n)'(\mathbf{x} - \bar{x}\mathbf{1}_n) = (\mathbf{M}^0\mathbf{x})'(\mathbf{M}^0\mathbf{x}) = \mathbf{x}'\mathbf{M}^{0\prime}\mathbf{M}^0\mathbf{x} = \mathbf{x}'\mathbf{M}^0\mathbf{x}$$
since $\mathbf{M}^0$ is symmetric and $\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$.

Idempotent Matrix

An idempotent matrix, **P**, is one that is equal to its square, that is, $\mathbf{P}^2 = \mathbf{PP} = \mathbf{P}$. It can be verified that if **P** is idempotent, then $(\mathbf{I} - \mathbf{P})$ is also an idempotent matrix. If **P** is a symmetric idempotent matrix (all of the idempotent matrices we shall encounter are symmetric), then $\mathbf{P}'\mathbf{P} = \mathbf{P}$. Thus, $\mathbf{M}^0$ is a symmetric idempotent matrix.

Consider constructing a matrix of sums of squares and cross products in deviations from the column means. For two vectors **x** and **y**,

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (\mathbf{M}^0\mathbf{x})'(\mathbf{M}^0\mathbf{y}) = \mathbf{x}'\mathbf{M}^0\mathbf{y}.$$

Hence, $\begin{bmatrix} \sum_{i=1}^{n}(x_i - \bar{x})^2 & \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) & \sum_{i=1}^{n}(y_i - \bar{y})^2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{M}^0\mathbf{x} & \mathbf{x}'\mathbf{M}^0\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{x} & \mathbf{y}'\mathbf{M}^0\mathbf{y} \end{bmatrix}$

Rank

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}$ are said to be linearly independent if the following equality holds

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0}$$

only when $\alpha_1 = \alpha_2 = \cdots\alpha_k = 0$. It is easy to see that no vector in a set of linearly independent vectors can be expressed as a linear combination of the other vectors.

Given $\mathbf{X}_{n\times k} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots.\ \mathbf{x}_k] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$,

if among the k columns $\mathbf{x}_j$ of **X** only c are linearly independent, then we say the column rank of **X** is c. If all k columns $\mathbf{x}_j$ of **X** are linearly independent, then we say **X** has full column rank. The row rank of **X** can be defined similarly. The smaller of the column rank and the row rank of **X** is referred to as the rank of **X** and denoted as rank(**X**). Given that $k < n$, i.e., **X** has more rows than columns, then rank(**X**) = k if X has full column rank. The implication of full column rank is that if **X** is not of full column rank, rank(**X**) < k, then at least one of its column is a linear combination of the other k – 1 columns.

<u>Some properties</u>:

1. $\text{Rank}(\mathbf{A}_{n\times k}) \le \min\{n, k\}$: The rank of a matrix cannot exceed its numbers of rows and columns.
2. When $\text{Rank}(\mathbf{A}_{n\times n}) = n$, **A** is nonsingular, i.e., $\mathbf{A}^{-1}$ exists.

3. When $\text{Rank}(\mathbf{A}_{n\times n}) < n$, then **A** is singular and $\mathbf{A}^{-1}$ does not exist.

4. Rank(**AB**) ≤ min{Rank(**A**), Rank(**B**)}

5. If **B** is a <u>square</u> matrix with <u>full</u> rank, Rank(**AB**) = Rank(**A**).

6. $\text{Rank}(\mathbf{X}'\mathbf{X}) = \text{Rank}(\mathbf{X}\mathbf{X}') = \text{Rank}(\mathbf{X})$.

Eigenvalues and Eigenvectors of Symmetric Matrices

A useful set of results fro analyzing a square matrix **A** arises from the solutions to the set of equations

$$\mathbf{A}\mathbf{u} = \lambda\,\mathbf{u}.$$

The pairs of solutions are the eigenvectors **u** and eigenvalues λ. If **u** is any solution vector, then k**u** is also for any value of k. To remove the indeterminacy, **u** is *<u>normalized</u>* so that $\mathbf{u}'\mathbf{u} = 1$. The solution then consists of λ and (n – 1) unknown elements in **u**.

The above set of equations can be rewritten as:

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{u} = \mathbf{0}.$$

If (**A** – λ**I**) is nonsingular, the only solution is **u** = **0**. Hence, the condition for **u** and λ exist (other than **u** = **0**) is that (**A** – λ**I**) is nonsingular. Hence, if λ is a solution, then

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

This equation is called the characteristic equation of **A**. For a symmetric matrix **A** of order n, its characteristic equation is an nth-order polynomial in λ. There are n real roots to be denoted $\lambda_1, \lambda_2, \ldots, \lambda_n$, some of which may be zero. Corresponding to each λ_i is a vector $\mathbf{u}_i$ satisfying the following equation:

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad \text{for} \quad i = 1, 2, \ldots, n.$$

Note that for symmetric matrix, eigenvectors $\mathbf{u}_i$'s are distinct (the corresponding eigenvalues λ_i's, although real, may not be distinct) and orthogonal, i.e., $\mathbf{u}_i'\mathbf{u}_j = 0$ for $i \neq j$. It is convenient to collect the n-eigenvectors in a $n\times n$ matrix whose ith column is the $\mathbf{u}_i$ corresponding λ_i,

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n],$$

and the n-eigenvalues in the same order, in a diagonal matrix,

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then, the full set of equations

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

is contained in

$$\mathbf{AU} = \mathbf{U}\Lambda .$$

Since eigenvectors are orthogonal $\mathbf{u}_i'\mathbf{u}_j = 0$ for $i \neq j$ and normalized $\mathbf{u}_i'\mathbf{u}_i = 1$, we have

$$\mathbf{U}'\mathbf{U} = \mathbf{I} .$$

This implies that $\mathbf{U}' = \mathbf{U}^{-1}$. It can be shown that $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}' = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i'$ (spectral decomposition of a matrix **A**) and $\mathbf{U}'\mathbf{A}\mathbf{U} = \boldsymbol{\Lambda}$ (diagonalization of a matrix **A**).

Example: $\mathbf{A} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$. Its characteristic equation is

$$\det \begin{bmatrix} 5-\lambda & -3 \\ -3 & 5-\lambda \end{bmatrix} = (5-\lambda)^2 - 9 = 0 .$$

Hence, we have $\lambda_1 = 2$ and $\lambda_2 = 8$; $\mathbf{u}_1 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$ and $\mathbf{u}_2 = [\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$.

<u>Some properties</u>:

1. $\det(\mathbf{A}) = \det(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}') = \det(\mathbf{U})\det(\boldsymbol{\Lambda})\det(\mathbf{U}') = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n} \lambda_i$.
2. tr(A) = $\text{tr}(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}') = \text{tr}(\boldsymbol{\Lambda}\mathbf{U}'\mathbf{U}) = \text{tr}(\boldsymbol{\Lambda}) = \sum_{i=1}^{n} \lambda_i$.
3. Since **U** is nonsingular, rank(**A**) = $\text{rank}(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}') = \text{rank}(\boldsymbol{\Lambda})$, the rank of **A** is equal to the number of nonzero eigenvalues.
4. The eigenvalues of a nonsingular matrix are all nonzero.
5. Let λ be an eigenvalue of **A**.
   (1) When **A** is singular, λ^k is an eigenvalue of $\mathbf{A}^k$ for <u>positive integer</u> k.

$$\mathbf{A}^2\mathbf{u} = \mathbf{A}\lambda\mathbf{u} = \lambda\mathbf{A}\mathbf{u} = \lambda(\lambda\mathbf{u}) = \lambda^2\mathbf{u} \quad \Rightarrow \quad \mathbf{A}^k\mathbf{u} = \lambda^k\mathbf{u}$$

   (2) When **A** is nonsingular, λ^k is an eigenvalue of $\mathbf{A}^k$ for <u>integer</u> k; especially, if **A** is nonsingular with eigenvalue λ, the inverse $\mathbf{A}^{-1}$ has λ^{-1} as an eigenvalue.

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \Rightarrow \mathbf{u} = \mathbf{A}^{-1}\lambda\mathbf{u} = \lambda\mathbf{A}^{-1}\mathbf{u} .$$

   Note that $\mathbf{A}^0 = \mathbf{I}$.

   (3) For a scalar a, $a\lambda$ is an eigenvalue of $a\mathbf{A}$ $(a\mathbf{A}\mathbf{u} = a\lambda\mathbf{u})$.
6. The eigenvectors of **A** and $\mathbf{A}^k$ are the same.
7. The eigenvalues of an idempotent matrix are either 0 or 1.
   (1) The only full rank, symmetric idempotent matrix is the identity matrix I.
   (2) All symmetric idempotent matrices except the identity matrix are singular.
8. The rank of an idempotent matrix is equal to its trace.

Quadratic Forms and Definite Matrices

Many optimization problems involve double sums of the form

$$q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \ .$$

This quadratic form can be written as:

$$q = \mathbf{x}'\mathbf{A}\mathbf{x} \, ,$$

where $\mathbf{A} = [a_{ij}]_{n \times n}$ is a symmetric matrix with order n and $\mathbf{x} \in \Re^n$. In general, q may be positive or negative, depending on **A** and **x**. However, there are some matrices for which q will be positive (or negative) regardless of **x**.

A symmetric matrix **A** with order n is:

1. Positive definite (p.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} > \mathbf{0}$ for all $\mathbf{x} \in \Re^n$, except $\mathbf{x} \neq \mathbf{0}$;
2. Negative definite (n.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} < \mathbf{0}$ for all $\mathbf{x} \in \Re^n$, except $\mathbf{x} \neq \mathbf{0}$;
3. Positive semi-definite (p.s.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq \mathbf{0}$ for all $\mathbf{x} \in \Re^n$;
4. Negative semi-definite (n.s.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} \leq \mathbf{0}$ for all $\mathbf{x} \in \Re^n$.

Some properties:

1. The eigenvalues of a p.d. (n.d.) matrix are all positive (negative).
2. The eigenvalues of a p.s.d. (n.s.d.) matrix are all positive or zero (negative or zero).
3. If **A** is p.d. (p.s.d.), then det(**A**) > (≥) 0.
4. If **A** is n.d. (n.s.d.), then det(**A**) < (≤) 0 for odd order and det(**A**) > (≥) 0 for even order.
5. If **A** is p.d. (n.d.), so is $\mathbf{A}^{-1}$.
6. The identity matrix **I** is p.d.
7. When **A** is p.d., λ^k is an eigenvalue of $\mathbf{A}^k$ for any real number k. Especially, $\mathbf{A}^{1/2} \equiv \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}'$ and $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$. Note that the eigenvalues of $\mathbf{A}^{1/2}$ are the square roots of the eigenvalues of **A**.
8. If **A** is $n \times K$ with full column rank and $n > K$, then $\mathbf{A}'\mathbf{A}$ is p.d. and $\mathbf{A}\mathbf{A}'$ is p.s.d.

   Since $\mathbf{A}\mathbf{x} \neq \mathbf{0}$ for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{y} = \sum_{i=1}^{K} y_i^2 > \mathbf{0}$.
9. Every idempotent matrix is p.s.d.
10. If **A** is symmetric and idempotent, $n \times n$ with rank J, then every quadratic form in **A** can be written as $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{J} y_i^2$
11. Suppose that both **A** and **B** have the same dimensions. We say that **A** is **larger** than **B** (**A** > **B**) if **A** – **B** is positive definite.

$$d = \mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}'(\mathbf{A} - \mathbf{B})\mathbf{x}$$

12. If **A** > **B**, then $\mathbf{B}^{-1} > \mathbf{A}^{-1}$.

Moments of Random Vectors

Definition: Let $\boldsymbol{y}$ be an $n \times 1$ random vector.

(1) The expected value of $\boldsymbol{y}$, denoted $E(\boldsymbol{y})$, is the vector of expected values:

$$E(\mathbf{y}) = \left[E(y_1), E(y_2), \ldots, E(y_n)\right]'.$$

(2) Its variance-covariance matrix, denoted Var($\boldsymbol{y}$), is defined as

$$\operatorname{Var}(\mathbf{y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

where $\sigma_j^2 = \operatorname{Var}(y_j)$ and $\sigma_{ij} = \operatorname{Cov}(y_i, y_j)$. It is obvious that a variance-covariance matrix is symmetric.

Properties: Let **A** be an $m \times n$ nonrandom matrix and **a** be an $n \times 1$ nonrandom vector.

(1) $E(\mathbf{A}\boldsymbol{y} + \mathbf{a}) = \mathbf{A}E(\boldsymbol{y}) + \mathbf{a}$.

(2) $\operatorname{Var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\operatorname{Var}(\mathbf{y})\mathbf{a} \geq 0$.

(3) If $\operatorname{Var}(\mathbf{a}'\mathbf{y}) > 0$ for all $\mathbf{a} \neq \mathbf{0}$, Var($\boldsymbol{y}$) is positive definite.

(4) $\operatorname{Var}(\mathbf{y}) = E\left[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\right]$, where $\boldsymbol{\mu} = E(\boldsymbol{y})$.

(5) If the elements are uncorrelated, $\sigma_{ij} = 0$ for $i \neq j$, Var($\boldsymbol{y}$) is a diagonal matrix. If, in addition, $\operatorname{Var}(y_j) = \sigma^2$ for $j = 1, 2, \ldots, n$, then $\operatorname{Var}(\boldsymbol{y}) = \sigma^2\mathbf{I}_n$.

(6) $\operatorname{Var}(\mathbf{A}\mathbf{y} + \mathbf{a}) = \mathbf{A}\left[\operatorname{Var}(\mathbf{y})\right]\mathbf{A}'$.

Distribution of Quadratic Forms

Let $\mathbf{y}$ be an $n \times 1$ multivariate normal random vector with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, written as $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Properties: Let **A** and **B** be square matrices with order n, and **b** be an $n \times 1$ vector.

(1) Each element of **y** is normally distributed.

(2) Any two elements of **y**, y_i and y_j, are independent if and only if they are uncorrelated, that is, $\sigma_{ij} = 0$.

(3) $\mathbf{Ay} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b},\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

(4) If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$, then $\mathbf{y}'\mathbf{y} \sim \chi_n^2$.

(5) If $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{I}_n)$

(6) If $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim \chi_n^2$.

(7) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$ and **A** is symmetric idempotent, then $\mathbf{y}'\mathbf{A}\mathbf{y} \sim \chi_q^2$ where q = tr(**A**).

(8) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$, then $\mathbf{y}'\mathbf{M}^0\mathbf{y} \sim \chi_{(n-1)}^2$.

(9) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$ and **A** and **B** are symmetric idempotent, then $\mathbf{y}'\mathbf{A}\mathbf{y}$ and $\mathbf{y}'\mathbf{B}\mathbf{y}$ are <u>independent</u> if **AB** = **0**.

(10) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$, **A** and **B** are symmetric idempotent, and **AB** = **0**, then

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}/r_a}{\mathbf{y}'\mathbf{B}\mathbf{y}/r_b} \sim F[r_a, r_b]$$

where $r_a = \text{tr}(\mathbf{A})$ and $r_b = \text{tr}(\mathbf{B})$.

(11) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$, **A** is symmetric idempotent, and **L** be an $m \times n$ matrix, then $\mathbf{Ly}$ and $\mathbf{y}'\mathbf{A}\mathbf{y}$ are <u>independent</u> if **LA** = **0**.

(12) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$ and **Ab** = **0**, $\dfrac{\mathbf{b}'\mathbf{y}/\sqrt{\mathbf{b}'\mathbf{b}}}{\sqrt{\mathbf{y}'\mathbf{A}\mathbf{y}/q}} \sim t_q$ where q = tr(**A**).

(13) If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_n)$, $\dfrac{(1/\sqrt{n})\ \mathbf{1}'\mathbf{y}}{\sqrt{\mathbf{y}'\mathbf{M}^0\mathbf{y}/(n-1)}} \sim t_{(n-1)}$.

Differential Operators

This section consider three kinds of differential calculus: Scalar, vectors, and quadratic forms.

1. **Scalars**: $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$. The symbol $\frac{\partial}{\partial \mathbf{x}}$ represent a whole vector of differential operators.

Example: $q = 3x_1 + 4x_2 + 9x_3 = [3 \quad 4 \quad 9]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{a}'\mathbf{x}$. $\frac{\partial q}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}$.

2. **Vectors**: $\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A} \quad \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}'$.

Give

$$\mathbf{y}' = [y_1 \quad y_2 \quad \cdots \quad y_n] = [\mathbf{x}'\mathbf{a}_1 \quad \mathbf{x}'\mathbf{a}_2 \quad \cdots \quad \mathbf{x}'\mathbf{a}_n] = \mathbf{x}'\mathbf{A}$$

where the ith element of **y** is $y_i = \mathbf{x}'\mathbf{a}_i$ for $\mathbf{a}_i$ being the ith column of **A**, with $i = 1,\ldots, n$.

$$\frac{\partial \mathbf{y}'}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} & \frac{\partial y_2}{\partial \mathbf{x}} & \cdots & \frac{\partial y_n}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{x}'\mathbf{a}_1}{\partial \mathbf{x}} & \frac{\partial \mathbf{x}'\mathbf{a}_2}{\partial \mathbf{x}} & \cdots & \frac{\partial \mathbf{x}'\mathbf{a}_n}{\partial \mathbf{x}} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n] = \mathbf{A}.$$

Example: $\mathbf{A}\mathbf{x} = \begin{bmatrix} 2 & 6 & -1 \\ 3 & -2 & 4 \\ 3 & 4 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 - x_3 \\ 3x_1 - 2x_2 + 4x_3 \\ 3x_1 + 4x_2 + 7x_3 \end{bmatrix}$ gives

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial(2x_1 + 6x_2 - x_3)}{\partial \mathbf{x}} & \frac{\partial(3x_1 - 2x_2 + 4x_3)}{\partial \mathbf{x}} & \frac{\partial(3x_1 + 4x_2 + 7x_3)}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 \\ 6 & -2 & 4 \\ -1 & 4 & 7 \end{bmatrix} = \mathbf{A}'$$

3. **Quadratic forms**: $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathbf{P}}{\partial \mathbf{x}} + \frac{\partial \mathbf{Q}\mathbf{x}}{\partial \mathbf{x}}$, where **P** = **Ax** and **Q** = $\mathbf{x}'\mathbf{A}$; hence,

$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{P} + \mathbf{Q}' = \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x}$. Furthermore, if **A** is symmetrical, then $\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$.

Example: $q = [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 1 & 3 & 5 \\ 3 & 4 & 7 \\ 5 & 7 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + 6x_1x_2 + 10x_1x_3 + 4x_2^2 + 14x_2x_3 + 9x_3^2$,

$$\frac{\partial q}{\partial \mathbf{x}} = \begin{bmatrix} \partial q/\partial x_1 \\ \partial q/\partial x_2 \\ \partial q/\partial x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 6x_2 + 10x_3 \\ 6x_1 + 8x_2 + 14x_3 \\ 10x_1 + 14x_2 + 18x_3 \end{bmatrix} = 2\mathbf{A}\mathbf{x}.$$

The Matrix Form of Simple Linear Regression Models

Consider the simple regression model with N observations

$$y_n = \beta_1 + \beta_2 x_n + \varepsilon_n, \qquad n = 1, 2, \ldots, N \tag{B1}$$

This can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{B2}$$

or

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{B3}$$

where $\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ $\quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$ $\quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ $\quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$

Assumptions of the linear Regression model:

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

(2) $E(\varepsilon) = \begin{bmatrix} E(\varepsilon_1) \\ \vdots \\ E(\varepsilon_N) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$

(3) $E(\varepsilon\,\varepsilon') = \begin{bmatrix} E(\varepsilon_1^2) & \cdots & E(\varepsilon_1\varepsilon_N) \\ \vdots & \ddots & \vdots \\ E(\varepsilon_N\varepsilon_1) & \cdots & E(\varepsilon_N^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_N.$

(4) $\mathbf{X}$ is an $N \times 2$ matrix with det $(\mathbf{X}'\,\mathbf{X}) \neq 0$.

(5) (Optional) $\quad \varepsilon \sim N(\mathbf{0},\ \sigma^2 \mathbf{I}_N).$

Under the assumptions discussed before, the best (minimum variance) linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is obtained by minimizing the error sum of squares

$$\mathbf{S}(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{B4}$$

This is known as the Gauss-Markov theorem.

Using the formulas for vector differentiation, we have

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2X'X\beta\,. \tag{B5}$$

Setting the equation (B6) to be zero gives the normal equation

$$(\mathbf{X}'\,\mathbf{X})\,\hat{\boldsymbol{\beta}} \;=\mathbf{X}'\,\mathbf{Y}. \tag{B6}$$

Since the square matrix ($\mathbf{X}'\,\mathbf{X}$) is non-singular, the OLS estimator $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} \;= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{B7}$$

Substituting equation (B3) into equation (B7), we get

$$\hat{\boldsymbol{\beta}} \;= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\,\mathbf{X}\beta + \varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

Since E($\boldsymbol{\varepsilon}$) = 0, we have $E\left(\hat{\beta}\right) = \beta$. Thus $\hat{\beta}$ is an unbiased estimator. Also,

$$\mathbf{Var}\left(\hat{\boldsymbol{\beta}}\right) \;=\; E\left(\hat{\boldsymbol{\beta}} - \beta\right)\left(\hat{\boldsymbol{\beta}} - \beta\right)' = \left(X'\,X\right)^{-1} X'\,E\left(\varepsilon\varepsilon'\right)X\left(X'\,X\right)^{-1}$$

$$= \sigma^2\left(X'\,X\right)^{-1} \quad \text{since} \quad E\left(\varepsilon\varepsilon'\right) = \sigma^2 I_N\,.$$

The $\hat{\beta}$ is unbiased and has a covariance matrix $\sigma^2\left(X'\,X\right)^{-1}$.

According to equation (B7), the vectors of fitted values $\hat{Y}$ and the least squares residuals $\hat{\varepsilon}$ are

$$\hat{Y} \;= \mathbf{X}\hat{\boldsymbol{\beta}} \;= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P_X Y} \tag{B?}$$

$$\hat{\varepsilon} \;= \mathbf{Y} - \;\hat{Y} = \mathbf{Y} - \mathbf{P_X Y} = (\mathbf{I} - \mathbf{P_X})\mathbf{Y} = \mathbf{M_X Y} \tag{B\$}$$

where $\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M_X} = (\mathbf{I}_N - \mathbf{P_X})$. **Note** that it can be shown that $\mathbf{P_X\,P_X = P_X}$ and $\mathbf{M_X\,M_X = M_X}$ (P_X and M_X are called **<u>idempotent</u>** matrices).

The total variation of the dependent variable is the sum of squared deviations from its mean (SST):

$$\mathbf{SST} = \;\textstyle\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2 \;= \mathbf{Y}'\,\mathbf{Y} - \;n\bar{y}^2\,.$$

It can be shown that total sum of squares = regression sum of squares + error sum of squares (**SST** = **SSR** + **SSE**), where

$$\mathbf{SSR = Y'P_XY} - n\bar{y}^2 \;= \mathbf{Y'X(X'X)^{-1}X'Y} - n\bar{y}^2 \;=\; \hat{\boldsymbol{\beta}}'X'Y - n\bar{y}^2$$

$$\mathbf{SSE} = \;Y'\left(I_N - P_X\right)Y \;=\; Y'M_XY \;=\; Y'M_X'M_XY \;=\; \hat{\varepsilon}'\hat{\varepsilon}$$

The unbiased estimator of σ^2 is

$$\hat{\boldsymbol{\sigma}}^2 = \frac{\boldsymbol{SSE}}{\boldsymbol{N}-2} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N-2} = \frac{\boldsymbol{Y}'(\boldsymbol{I}_N - \boldsymbol{P}_X)\boldsymbol{Y}}{N-2}.$$

We now calculate elements of some matrices discussed above in order to verify some results derived in earlier classes:

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} = \begin{bmatrix} N & \sum_{n=1}^N x_n \\ \sum_{n=1}^N x_n & \sum_{n=1}^N x_n^2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N x_n y_n \end{bmatrix}.$$

Base on the above results, we have

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix} \begin{bmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N x_n y_n \end{bmatrix}$$

$$= \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 \sum_{n=1}^N y_n - \sum_{n=1}^N x_n \sum_{n=1}^N x_n y_n \\ N\sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n \sum_{n=1}^N y_n \end{bmatrix}$$

The covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2\left(X'X\right)^{-1} = \frac{\sigma^2}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix}$$

The Matrix Form of Multiple Linear Regression Models

Suppose that we have the following N observations:

$$
\begin{array}{ccccccccccccc}
y_1 & = & \beta_1 & + & \beta_2 x_{12} & + & \beta_3 x_{13} & + & \cdots & + & \beta_K x_{1K} & + & \varepsilon_1 \\
y_2 & = & \beta_1 & + & \beta_2 x_{22} & + & \beta_3 x_{23} & + & \cdots & + & \beta_K x_{2K} & + & \varepsilon_2 \\
\vdots & & \vdots & & \vdots & & \vdots & & \cdots & & \vdots & & \vdots \\
y_N & = & \beta_1 & + & \beta_2 x_{N2} & + & \beta_3 x_{K3} & + & \cdots & + & \beta_K x_{NK} & + & \varepsilon_N
\end{array}
$$

The matrix form is

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

where

$$
\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad
\boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1K} \\ 1 & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & \cdots & x_{NK} \end{bmatrix} \qquad
\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \qquad
\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}
$$

The assumptions of the multiple linear Regression model are the same as the simple linear regression model *except* $\mathbf{X}$ being an $N \times K$ matrix. Other matrix algebras are exactly the same as the simple linear regression model