Definition: An $m \times n$ matrix, $\mathbf{A}_{m\times n}$, is a rectangular array of real numbers with m rows and n columns. Element in the i^{th} row and the j^{th} column is denoted by a_{ij}.

$$\mathbf{A}_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{bmatrix}_{m\times n}$$

Definition: A vector **a** of length n is an $n \times 1$ matrix with each element denoted by a_i. The i^{th} element is called the i^{th} component of the vector and n is the dimensionality.

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Matrix Operations

1. Two matrices **A** and **B** of the same dimensions can be added. The sum **A** + **B** has (i, j) entry $a_{ij} + b_{ij}$. So

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}$$

$$\text{Example : } \mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \\ -1 & -3 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 2 \\ 9 & 5 \\ 3 & 0 \end{bmatrix} \quad \mathbf{A} + \mathbf{B} = \begin{bmatrix} 4 & 5 \\ 11 & 11 \\ 2 & -3 \end{bmatrix}$$

   - **A + B = B + A**
   - **(A + B) + C = A + (B + C)**

2. A matrix may also be multiplied by a constant c. The product c**A** is the matrix that results from multiplying each element of **A** by c. Thus

$$c\mathbf{A}_{m\times n} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{21} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & \cdots & \cdots & ca_{mn} \end{bmatrix}_{m\times n}$$

3. The **transpose** operation $\mathbf{A}^T$ or $\mathbf{A}'$ of a matrix changes the columns into rows so that the first column of $\mathbf{A}$ becomes the first row of $\mathbf{A}^T$, the second column becomes second row, and etc. So the (i,j)$^{\text{th}}$ element in $\mathbf{A}_{m\times n}$ becomes the (j,i)$^{\text{th}}$ in the transpose $\mathbf{A}^T_{n\times m}$.

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

- $(\mathbf{A}^T)^T = A$
- $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$
- $(c\cdot \mathbf{A})^T = c\cdot \mathbf{A}^T$

4. We can define matrix multiplication **A B** if the number of elements in a row of A is the same as the number of elements in the columns of B. E.g. when **A** is (p $\times$ k) and **B** is (k $\times$ n). An element of the new matrix AB is formed by taking the inner product of each row of A with each column of B. The matrix product AB is

$$\mathbf{A}_{m\times n} \times \mathbf{B}_{n\times m} = \mathbf{C}_{m\times n} \text{ with } c_{ij} = \sum_{k=1}^{n} a_{ik} + b_{kj}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}\cdot b_{11} + a_{12}\cdot b_{21} & a_{11}\cdot b_{12} + a_{12}\cdot b_{22} \\ a_{21}\cdot b_{11} + a_{22}\cdot b_{21} & a_{21}\cdot b_{12} + a_{22}\cdot b_{22} \end{bmatrix}$$

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

$$\mathbf{A}\,\mathbf{A}^T = \begin{bmatrix} 1(1)+2(2)+3(3) & 1(4)+2(5)+3(6) \\ 4(1)+5(2)+6(3) & 4(4)+5(5)+6(6) \end{bmatrix}_{2\times 2}$$

- **(AB)C = A(BC)**
- **A(B+C) = AB+AC**
- **(A+B)C = AC+BC**
- **AB** does not generally equal to **BA**!

Special Square Matrix

1. A square matrix **A** is said to be **symmetric** if **A** = $\mathbf{A}^T$ or $a_{ij} = a_{ji}$ for all i and j.

$$\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 5 & -2 \end{bmatrix}_{2\times 2} \text{ is symmetric;} \qquad \mathbf{A} = \begin{bmatrix} 3 & 5 \\ 6 & -2 \end{bmatrix}_{2\times 2} \text{ is not symmetric}$$

2. Diagonal Matrix: $\mathbf{A}_{n\times n} = \text{diag}(a_1, \ldots, a_n) = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & a_n \end{pmatrix}$

3. Identity Matrix: $\mathbf{I}_{n\times n} = \text{diag}(\mathbf{1}_n) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$

   The identity matrix is a square matrix with ones on the diagonal and zeros elsewhere. It follows from the definition of matrix multiplication that the (i, j) enrty of **AI** is $a_{i1} \times 0 + \ldots + a_{i,j1} \times 0 + a_{ij} \times 1 + a_{i,j+1} \times 0 + \ldots + a_{ik} \times 0 = a_{ij}$. So **AI = A** . Similarly, **IA = A** . Therefore matrix **I** acts like 1 in ordinary multiplication.

   The fundamental scalar relation about the existence of an inverse number a^{-1} such that $a^{-1} \cdot a = a \cdot a^{-1} = 1$, if $a \neq 0$, has the following matrix algebra extension.

   $$\mathbf{B}_{\,k\times k}\,\mathbf{A}_{\,k\times k} = \mathbf{A}_{\,k\times k}\,\mathbf{B}_{\,k\times k} = \mathbf{I}_{k\times k}$$

   then **B** is called the inverse of **A** and is denoted by A^{-1}.

Other Matrix Properties

1. **Trace**: The sum of the diagonal elements, $tr(\mathbf{A}_{n\times n}) = \sum_{i=1}^{n} a_{ii}$.

2. A square matrix that does not have a matrix inverse is called a **singular** matrix
   The inverse of a 2×2 matrix is given by

   $$\mathbf{A}_{\,2\times 2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \mathbf{A}_{\,2\times 2}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

3. A matrix is singular if and only if its determinant is 0. The determinant of a matrix **A** is denoted as $|\mathbf{A}\,|$. The determinant of a 2×2 matrix is given by

   $$\mathbf{A}_{\,2\times 2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad det(A) = ad - bc$$

Examples 1: Simultaneous equations

$$\begin{aligned} 5x + 3y + z &= 1 \\ 2x + 3y + 5z &= 2 \\ x + 9y + 6z &= 3 \end{aligned}$$

We can rewrite the above three equations as a single matrix equation:

$$\begin{bmatrix} 5 & 3 & 1 \\ 2 & 3 & 5 \\ 1 & 9 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Example 2: Variance/Covariance Matrix

For a vector of random variables, $(Y_1, Y_2, \ldots, Y_n)$, we can write a matrix containing their variances and their covariances. Let σ_i^2 be the variance of Y_i and let cov_{ij} be the covariance between Y_i and Y_j, $i < j$. Then the variance/covariance matrix for $(Y_1, Y_2, \ldots, Y_n)$ is

$$\begin{bmatrix} \sigma_1^2 & cov_{12} & \ldots & cov_{1n} \\ cov_{12} & \sigma_2^2 & \ldots & cov_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ cov_{1n} & cov_{2n} & \ldots & \sigma_n^2 \end{bmatrix}$$

Also note that the above can be written as

$$\begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{12} & 1 & \ldots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \ldots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{bmatrix}$$

where ρ_{ij} is the correlation of Y_i and Y_j. Note that all of these matrices are symmetric. Furthermore, the terms on the diagonal of the variance/covariance matrix must be positive and terms off the diagonal of the correlation matrix are bounded by -1 and 1.

Example 3: Multiple Linear Regression

We have a response Y and a set of p independnet variables $X_1, X_2, \ldots, X_p$. Assume we have n observations and for each $i = 1, \ldots, n$ observation, we assume

$$Y_i = \beta_0 + \beta_1 \, x_{i,1} + \ldots + \beta_p \, x_{i,p} + \varepsilon_i$$

where $\varepsilon_i \sim$ Normal $(0, \sigma^2)$.

We can represent n observations simultaneously in matrix form as

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p+1}\,\boldsymbol{\beta}_{p+1\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The residual $Y_i - (\beta_0 + \beta_1\,x_{i,1} + \ldots + \beta_p\,x_{i,p})$ can be written as

$$\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) = \begin{bmatrix} y_1 - (\beta_0 + \beta_1\,x_{1,1} + \beta_2\,x_{1,2} + \ldots + \beta_p\,x_{1,p}) \\ y_2 - (\beta_0 + \beta_1\,x_{2,1} + \beta_2\,x_{2,2} + \ldots + \beta_p\,x_{2,p}) \\ \vdots \\ y_n - (\beta_0 + \beta_1\,x_{n,1} + \beta_2\,x_{n,2} + \ldots + \beta_p\,x_{n,p}) \end{bmatrix}_{n\times 1}$$

The residual sum of squares (RSS) is

$$\begin{aligned} \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)^T\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) &= \begin{bmatrix} y_1 - \beta_0 - \beta_1\,x_{1,1} - \beta_2\,x_{1,2} - \ldots - \beta_p\,x_{1,p} \\ y_2 - \beta_0 - \beta_1\,x_{2,1} - \beta_2\,x_{2,2} - \ldots - \beta_p\,x_{2,p} \\ \vdots \\ y_n - \beta_0 - \beta_1\,x_{n,1} - \beta_2\,x_{n,2} - \ldots - \beta_p\,x_{n,p} \end{bmatrix}_{n\times 1}^T \\ &\times \begin{bmatrix} y_1 - \beta_0 - \beta_1\,x_{1,1} - \beta_2\,x_{1,2} - \ldots - \beta_p\,x_{1,p} \\ y_2 - \beta_0 - \beta_1\,x_{2,1} - \beta_2\,x_{2,2} - \ldots - \beta_p\,x_{2,p} \\ \vdots \\ y_n - \beta_0 - \beta_1\,x_{n,1} - \beta_2\,x_{n,2} - \ldots - \beta_p\,x_{n,p} \end{bmatrix}_{n\times 1} \\ &= \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1\,x_{i,1} - \ldots - \beta_p\,x_{i,p})^2 \end{aligned}$$

Minimizing the above RSS gives the usual least-squared estimate for β

$$\hat{\boldsymbol{\beta}}_{OLS} = \text{argmin}_{\boldsymbol{\beta}}\,\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)^T\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$