Chapter 1

Systems of Linear Equations

人工智慧 AI，基本上是要處理大量的資訊。這需要運用線性代數處理大量的多元一次的聯立方程組（Systems of Linear Equations）以及最佳化的問題（Optimization）。這一章要探討的是多元一次的聯立方程組。我們依然利用大家熟悉的加減消去法（或高斯消去法）來處理這類方程組。不過我們不再只關心如何解特定的聯立方程組，而會更著重於有系統地探討一般聯立方程組解的情況的理論。我們會用矩陣來表示一個聯立方程組，不過這裡的矩陣僅是為了方便起見而使用，不會涉及矩陣的性質。至於真正矩陣的運算及性質，我們留待下一章再詳述。

1.1. 一次聯立方程組及基本列運算

所謂 n 元一次的方程式就是有 n 個未知數（variable）的一次方程式（linear equation）。例如 $2x_1+5x_2-x_3+x_4=1$ 就是一個 4 元一次的聯立方程組（當然也可看成是 5 元或更多元）。n 元一次的方程式抽象的表示法就是

$$a_1x_1+\cdots+a_nx_n=b,$$

其中這些 $a_1,\ldots,a_n$ 和 b 都是實數，而這些 x_i 表未知數。當我們有多個 n 元一次的方程式要討論它們的共同解時，就稱為解一次聯立方程組（system of linear equations）。一般抽象的表示法

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

表示有 m 個 n 元一次方程式所成的方程組（system of m linear equations in n variables）。這裡 $a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1$ 表示第一個方程式、$a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2$ 表示第二個方程式、而當 $1\leq i\leq m$ 時，第 i 個方程式就是 $a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n=b_i$，所以

最後一個（即第 m 個）方程式就是 $a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$。這裡 a_{ij}, b_i 皆為實數，這些實數才是真正影響到聯立方程組的因素，所以我們也可特別把它們標明出來。令

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

然後將上面的聯立方程組用 $A\mathbf{x} = \mathbf{b}$ 來表示。矩陣 A 中的每一個 a_{ij} 稱為 A 的一個 *entry*。因為 A 的每一個 entry 對應到聯立方程組中某個未知數的係數，通常我們會稱矩陣 A 為此聯立方程式的係數矩陣。一個矩陣的一個橫排稱為一個 row（列）；而一個豎排稱為一個 column（行）。我們算 row 時是從上而下來數的，也就是說最上面的一個 row 稱為第一個 row、下一個 row 稱為第二個 row、依此類推。而算 column 是由左而右來數的，也就是說最左邊的一個 column 稱為第一個 column、再往右一個 column 稱為第二個 column、依此類推。大家可以看出矩陣 A 的 row 對應的就是此聯立方程組的方程式：第一個 row 對應到第一個方程式、第二個 row 對應到第二個方程式、依此類推。而 column 對應到的是方程組的未知數：第一個 column 對應到的是未知數 x_1 的係數、第二個 column 對應到的是未知數 x_2 的係數、依此類推。因為這裡是由 m 個方程式而且每個方程式有 n 個未知數所組成的聯立方程組，所以 A 共有 m 個 row 以及 n 個 column。我們稱這樣的矩陣為 $m \times n$ matrix。注意這裡 $\mathbf{x}$ 表示是一個未知的向量而且我們將向量 $\mathbf{x}, \mathbf{b}$ 都寫成 column vector（行向量）是為了配合將來矩陣乘法的寫法。目前大家只要記住這也是聯立方程式的一種表示法即可。

例如解聯立方程組

$$\begin{array}{rcl} 3x_1 - 2x_2 + 9x_4 & = & 4 \\ 2x_1 + 2x_2 - 4x_4 & = & 6 \end{array} \tag{1.1}$$

我們就可以表成

$$\begin{bmatrix} 3 & -2 & 0 & 9 \\ 2 & 2 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}.$$

注意這裡係數矩陣多出 $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 這個 column 因為 x_3 的係數為 0。

過去學習解一次聯立方程組的方法不外加減消去法或高斯消去法，它們的原理都是一樣的，即利用以下三種基本方法：

(1) 變換式子的順序

(2) 將某一式乘上一非零實數

(3) 將某一式乘上一實數後加到另一式上

利用這三種基本方法將方程式的某些變數消去，最後求出解來。我們將介紹一個有系統的方法來解聯立方程組，把這三種基本方法看成是對矩陣的運算。

當我們要解

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

這一個聯立方程組時，先寫出如下的 augmented matrix （增廣矩陣）

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

例如式子 (1.1) 中的聯立方程組所對應的 augmented matrix 為

$$\left[\begin{array}{cccc|c} 3 & -2 & 0 & 9 & 4 \\ 2 & 2 & 0 & -4 & 6 \end{array}\right].$$

換言之，若我們要解 $A\mathbf{x} = \mathbf{b}$ 這一個聯立方程組，就要寫下 $[A \mid \mathbf{b}]$ 這一個 matrix。反之一個 augmented matrix $[A \mid \mathbf{b}]$ 就對應到一個聯立方程組 $A\mathbf{x} = \mathbf{b}$。

接下來我們將如加減消去法的三種步驟，利用所謂的 elementary row operation（基本列運算）處理這個 augmented matrix。所謂 elementary row operation 即表示對矩陣進行如下三種的列運算：

(1) 將矩陣的某兩個 row 對調

(2) 將矩陣的某一個 row 乘上一非零實數

(3) 將矩陣的某一個 row 乘上一實數後加到另一個 row

為了方便起見，我們將上面 (1)、(2)、(3) 三種 elementary row operation 分別稱為 *type* 1、*type* 2 以及 *type* 3 的 elementary row operation。

Example 1.1.1. 考慮矩陣

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & -1 & 3 \\ 4 & 0 & 1 & 2 \end{bmatrix}.$$

將 A 的第一、第二兩個 row 互換（即做一個 type 1 的 elementary row operation），可得

$$B = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 1 & 2 & 3 & 4 \\ 4 & 0 & 1 & 2 \end{bmatrix}.$$

而將 B 的第二個 row 乘上 2（即做一個 type 2 的 elementary row operation），可得

$$C = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 2 & 4 & 6 & 8 \\ 4 & 0 & 1 & 2 \end{bmatrix}.$$

而將 C 的第三個 row 乘上 -3 加到第一個 row（即做一個 type 3 的 elementary row operation），可得

$$D = \begin{bmatrix} -10 & 1 & -4 & -3 \\ 2 & 4 & 6 & 8 \\ 4 & 0 & 1 & 2 \end{bmatrix}. \quad \sharp$$

注意：若一個矩陣 P 經由一個 elementary row operation 轉換成矩陣 Q，我們也可以對 Q 藉由同樣 type 的 elementary row operation 將之轉換回 P。例如前面 Example 1.1.1 中，我們可以將 B 的第一、第二兩個 row 互換而得到 A。我們也可將 C 的第二個 row 乘上 1/2 而得到 B。另外我們也可將 D 的第三個 row 乘上 3 加到第一個 row 而轉換回 C（參見習題）。

Question 1.1. 兩個同樣 *type* 的 *elementary row operation* 交換順序會一樣嗎？兩個不同 *type* 的 *elementary row operation* 交換順序會一樣嗎？

1.2. 解聯立方程組

大家應很容易看出一個 augmented matrix 經過上節所提的 elementary row operation 後所得的 augmented matrix 所對應的聯立方程組就是大家熟悉的加減消去法的三種步驟所得的方程組。利用加減消去法最常遇到的問題是（尤其在處理未知數很多的方程組時）常常做了幾次後，混亂到不知道那些式子是消過了以及那些式子還可以再進一步消減；還有就是，到底要將方程組的式子消到哪種地步時，才可以解出方程組。關於第一個問題，我們可以理解用矩陣的表示法就可以把這消去的過程記錄下來。而接下來我們要探討的就是第二個問題，也就是將矩陣化成某種特定的形式就可以解出方程式來。

我們的目的就是要將 augmented matrix $[A \mid \mathbf{b}]$ 中的係數矩陣 A 利用這三種 elementary row operation 化成所謂的 *echelon form*。

我們先解釋一下何謂 echelon form。首先我們將矩陣每一個 row 從左到右來看第一個不為 0 的項稱為這個 row 的 *leading entry*。因為係數矩陣中的每一個 entry 對應到聯立方程組中某個 variable（未知數）的係數，所以 leading entry 若是 variable x_i 的係數，我們就說這個 leading entry 發生在 x_i 的位置。要注意，這也等同於這個 leading entry 是位於從左到右算來第 i 個 column。例如矩陣

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 4 \\ 0 & 0 & 5 & 0 & 2 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$$

第一個 row 的 leading entry 為 1 不過因為第一個 row 還有其他位置也是 1，所以我們特別要說明第一個 row 的 leading entry 發生在 x_1 的位置；而第二個 row 和第三個 row 的 leading entry 分別為 5 和 1 且發生的位置皆在 x_3。

所謂一個矩陣是 echelon form 表示這個矩陣沒有 leading entry 的 row（即該 row 每一項皆為 0）必需在最下方；而有 leading entry 的 row 其 leading entry 所在位置**從上到下來看是往右移**的。換言之，若上一個 row 的 leading entry 所在的位置是 x_i，而下一個 row 的 lading entry 所在的位置是 x_j，則必需 $i < j$。例如上一個矩陣並非 echelon form，因為第 3 個 row 和第 2 個 row 的 leading entry 的位置皆為 x_3，並未右移。另外矩陣

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & -1 \\ 3 & 0 & 0 & 0 \end{bmatrix}$$

都不是 echelon form。因為前一個矩陣全為 0 的 row 並未置於最下方，而後一個矩陣第 3 個 row 的 leading entry 在第 2 個 row 的 leading entry 的左方。至於矩陣

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

就是 echelon form。當一個矩陣是 echelon form 時，我們稱每一個 row 的 leading entry 為 *pivot*，而 pivot 所在的位置我們稱為 *pivot variable*。例如上一個矩陣其 pivot variables 依序為 x_2, x_3, x_5。

我們要強調，絕不會有 pivot 的個數多於方程組 variables（未知數）的個數的情形發生。這是因為當係數矩陣 A 是 echelon form 時，每一個 column 最多僅能有一個 pivot（因為不能有兩個 leading entry 在同一個 variable），所以 pivot 的個數不能多於 column 的個數。而 A 的 column 個數表示的就是此聯立方程組 variables 的個數，因此 pivot 的個數不會多於 variables 的個數。另一方面依定義每一個 row 最多僅能有一個 pivot，所以 pivot 的個數也不會多於該方程組的方程式個數（即係數矩陣 row 的個數）。

很容易就可以發現一個矩陣利用 elementary row operations 化成的 echelon form 的方法有很多種，而所化成的 echelon form 也有可能不同（例如一個 echelon form 的某一個 row 經由 type 2 的 elementary row operation 作用後仍為 echelon form）。不過以後我們會說明，同一個矩陣不管用哪些 elementary row operations 化成的 echelon form 它們的 pivot variables 皆會相同。也因此一個矩陣化為 echelon form 其 pivot 的個數是固定的，我們特別有以下的定義：

Definition 1.2.1. 假設 A 為一矩陣。若 A 利用 elementary row operations 化為 echelon form 後其 pivot 的個數為 r，我們稱 r 為 A 的 *rank*。用 $\mathrm{rank}(A) = r$ 來表示。

一個矩陣的 rank 是該矩陣很重要的資訊，以後我們會探討它很多的性質。由於它的重要性，我們特別在這裡先介紹它的定義，讓大家先熟悉一下。利用前面探討 pivot 的個數和 column、row 的個數之間的關係，由 rank 的定義知：如果矩陣 A 有 m 個 row 以及 n 個 column，則 $\mathrm{rank}(A) \leq \min\{m, n\}$。

Exercise 1.1. 假設 A, B 為同階的矩陣。以下各小題中 A 經由一個 elementary row operation 變成 B。請說明如何用同樣 type 的 elementary row operation 將 B 變回 A。

(1) A 經由將 i-th row 和 j-th row 交換的 type 1 elementary row operation 變成 B。

(2) A 經由將 i-th row 乘上非 0 實數 r 的 type 2 elementary row operation 變成 B。

(3) A 經由將 i-th row 乘上實數 r 加到 j-th row 的 type 3 elementary row operation 變成 B。

Exercise 1.2. 已知矩陣 A 經由 2nd 和 3rd rows 交換的 type 1 的 elementary row operation 變成 A_1，再將 A_1 經由 1st row 乘上 -2 加到 2nd row 的 type 3 的 elementary row operation 變成 A_2，最後 A_2 經由 3rd row 乘上 $1/2$ 的 type 2 的 elementary row operation 變成 $\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$。試依序寫下 A_2、A_1 以及 A。

Exercise 1.3. 考慮增廣矩陣

$$[A|\mathbf{b}] = \left[\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 2 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & -3 & -9 \end{array}\right]$$

(1) 請寫下增廣矩陣 $[A|\mathbf{b}]$ 所對應的聯立方程組。

(2) 請寫下係數矩陣 A 的 pivot variables、free variables 以及 $\operatorname{rank}(A)$。

Exercise 1.4. 令 $\lambda \in \mathbb{R}$, 考慮矩陣

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 4 \\ 1 & \lambda & 1 & 1 & 4 \\ 1 & 1 & \lambda & 3-\lambda & 6 \\ 2 & 2 & 2 & \lambda & 6 \end{bmatrix}$$

試依照 λ 的取值探討 $\operatorname{rank}(A)$。

Exercise 1.5. 考慮 type 1, type 2 以及 type 3 的 elementary row operations。

(1) 試著利用一些 type 3 以及 type 2 的 elementary row operations 將 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 變換成 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

(2) 證明所有 type 1 的 elementary row operation 皆可經由三次 type 3 的 elementary row operation 與一次 type 2 的 elementary row operation 得到。

(3) 利用 Exercise 1.1 之 elementary row operation 的可逆性以及前一小題說明某一個 row 乘上 -1 這種 type 2 的 elementary row operation，也可經由三次 type 3 的 elementary row operation 與一次 type 1 的 elementary row operation 得到。

(4) 說明當 $a \neq \pm 1$ 時，無法經一些 type 1 以及 type 3 的 elementary row operations 將 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 變換成 $\begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}$（Hint: 觀察行列式的變化）.（註: 這說明當 $a \neq \pm 1$ 時，某一個 row 乘上 a 這種 type 2 的 elementary row operation 無法經由一些 type 1 和 type 3 的 elementary row operations 得到。)

(5) 說明無法經一些 type 1 以及 type 2 的 elementary row operations 將 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 變換成 $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$（Hint: 觀察 0 的個數變化）.（註: 這說明 type 3 的 elementary row operation 無法經由一些 type 1 和 type 2 的 elementary row operations 得到。)