2

Matrices

Certainly the best times were when I was alone with mathematics, free of ambition and pretense, and indifferent to the world.

Robert Langlands

2.1 Matrix basics

Matrices are a useful abstractions of the notion of rectangular tables with numerical entries.

We will represent a matrix A with n rows and m columns like this:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

Note that each entry has two subindices: a_{ij} represents the entry that sits in the i-th row and j-th column.

We will denote the set of all matrices with n rows and m columns as $n \times m$ matrices or $\mathbb{R}^{n\times m}$.

A matrix is square if it has the same number of rows as columns.

2.2 Operations

Sum: two $n \times m$ matrices A and B can be summed together to give another $n \times m$ matrix C, doing the sum component by component:

$$A + B = C$$

where $c_{ij} = a_{ij} + b_{ij}$

Scaling: given a scaling factor $\alpha \in \mathbb{R}$ and an $n \times m$ matrix A, we can combine them to give another $n \times m$ matrix B:

$$\alpha A = B$$

where $b_{ij} = \alpha a_{ij}$

Multiplication: matrices can be combined in yet another way that we call *matrix multiplication* or *product*. There are two odd things about this operation that you must be aware of:

- The order of multiplication is important.

- Two matrices A, B can be multiplied together – with A coming first in the multiplication order, B coming second – as long as the number of columns of A coincides with the number of rows of B.

2.3 How matrix multiplication works

We present four ways of computing matrix multiplication that are equivalent, the reason of presenting several ways being that in some contexts it may be much easier to think in one way rather than any other.

Can you figure out why these four ways are equivalent to one another?

For the rest of the section, we will set a $n \times k$ matrix $A = (a_{ij})$ and a $k \times m$ matix $B = (b_{ij})$ that we want to multiply together to produce a matrix $C = AB = (c_{ij})$. Also it will be convenient to denote A_t the columns of A and B_t the rows of B, more graphically:

$$A = [A_1 \mid A_2 \mid \ldots \mid A_k]$$

and

$$B = \begin{bmatrix} B_1 \\ \hline B_2 \\ \hline \vdots \\ \hline B_k \end{bmatrix}$$

2.3.1 Rows versus columns

$$c_{ij} = \sum_{t=1}^{k} a_{it} b_{tj}$$

2.3.2 Columns

We can think of the result C column by column

$$C = [C_1 \mid C_2 \mid \dots \mid C_m]$$

where each column is obtained as

$$C_j = \sum_{t=1}^{k} b_{tj} A_t.$$

In other words, the columns of C are expressed as linear combinations of the columns of A.

2.3.3 Rows

We can think of the result C row by row

$$C = \begin{bmatrix} C_1 \\ \hline C_2 \\ \hline \vdots \\ \hline C_n \end{bmatrix}$$

where each row is obtained as

$$C_i = \sum_{t=1}^{k} a_{it} B_t$$

In other words, the rows of C are expressed as linear combinations of the rows of B.

2.3.4 Columns versus rows

Finally, we can think of the result C as being the sum of matrices that result from multiplying the columns of A with the rows of B – remember that A has as many columns as B has rows:

$$C = A_1B_1 + A_2B_2 + \ldots + A_kB_k.$$

2.4 Transpose

Given an $n \times m$ matrix A, we define its *transpose*, denoted A^t, as the $m \times n$ matrix whose rows are the columns of A and whose columns are the rows of A. Formally,

$$A^t = (a^t_{ij})$$

where $a^t_{ij} = a_{ji}$.

When a square matrix A is the same as its transpose A^t we say that A is *symmetric*.

2.5 Column space

Given a matrix A we will define the *column space* of A – denoted $C(A)$ – to be the vector space generated by the columns of A considered as vectors. In other words,

$$A = [c_1 \mid c_2 \mid \ldots \mid c_m]$$

then

$$C(A) = \mathbf{span}(c_1, c_2, \ldots, c_m)$$

2.6 Row space

Given a matrix A we will define the *row space* of A – denoted $R(A)$ – to be the vector space generated by the rows of A considered as vectors. In other words,

$$A = \left[\begin{array}{c} r_1 \\ \hline r_2 \\ \hline \vdots \\ \hline r_n \end{array}\right]$$

then

$$R(A) = \mathbf{span}(r_1, r_2, \ldots, r_n)$$

Note:

$R(A^t) = C(A)$ and $C(A^t) = R(A)$.

2.7 Rank

The vector spaces $C(A)$ and $R(A)$ may be very different. For one, for a $n \times m$ matrix, $C(A)$ is a vector subspace of $\mathbb{R}^n$, whereas $R(A)$ is a vector subspace of $\mathbb{R}^m$. However, they have something in common:

Proposition 5. *Given a matrix A, $\boldsymbol{dim}(C(A)) = \boldsymbol{dim}(R(A))$.*

To prove this important result we will put to work two of the ways to look at matrix multiplication: by columns and by rows, respectively. Assume A is an $n \times m$ matrix and $\mathscr{V} =$

$v_1, \dots, v_k$ is a basis of $C(A)$, where $k = \mathbf{dim}(C(A))$. Since each column of A can be expressed as a linear combination of the elements of $\mathscr{V}$, we can express the matrix A as a product $A = C\Lambda$ of two matrices, where

$$C = [v_1 \mid \dots \mid v_k]$$

and $\Lambda = (\lambda_{ij})$ is a $k \times m$ satisfying

$$A_j = \lambda_{1j} v_1 + \dots + \lambda_{kj} v_k$$

for every $j = 1, \dots, m$ (the column perspective of matrix multiplication). If we look back at the identity $A = C\Lambda$, this time using the row perspective of matrix multiplication, we can see that every row of A can be expressed as a linear combination of the k row vectors of the matrix Λ. In other words,

$$\mathbf{dim}(R(A)) \leq k = \mathbf{dim}(C(A)).$$

Since this inequality is completely general, we can now use it with A^t, the transpose of A, leading to

$$\mathbf{dim}(R(A^t)) \leq \mathbf{dim}(C(A^t))$$

but since $R(A^t) = C(A)$ and $C(A^t) = R(A)$, this means that also

$$\mathbf{dim}(R(A)) \geq \mathbf{dim}(C(A)).$$

The *rank* of a matrix A is defined as $\mathbf{dim}(C(A))$ or, equivalently, $\mathbf{dim}(R(A))$. We denote it as $\mathbf{rank}(A)$.

2.7.1 Remark

You can think about the rank as counting the maximum size of linearly independent sets made of rows (or columns) of A.

2.8 Rank 1 matrices

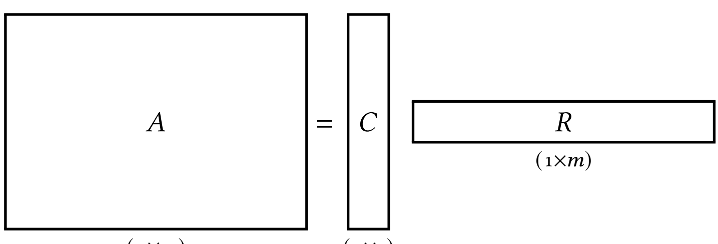
What happens when we multiply a 1-column matrix times a 1-row matrix? If B is an $n \times 1$ matrix, and C is a $1 \times m$ matrix, the product $A = BC$ is an $n \times m$ matrix. What is $\mathbf{rank}(A)$?

If we adopt the column perspective, A has as columns linear combinations of the only column that B has. In other words, all the columns of A are rescaled versions of the only column of B. We can then conclude that $\mathbf{rank}(A) = 1$.

It turns out that all rank 1 matrices can factorize as a product of a single column and a single row matrices.

Proposition 6. *For any $n \times m$ matrix A, $\mathbf{rank}(A) = 1$ if and only if there are matrices $B \in \mathbb{R}^{n\times 1}$ and $C \in \mathbb{R}^{1\times m}$ such that $A = BC$.*

In other words, all rank 1 matrices can be expressed as a product of *spaghetti* matrices like this:



2.9 Matrices define linear maps

Equipped with the idea of matrix multiplication, matrices define a particularly interesting type of numerical function known as *linear maps*.

A linear map takes a vector and transforms it into another vector, although not necessarily of the same size.

If we think of an m-vector v as a single column matrix and A is an $n \times m$ matrix, we can then multiply Av giving another single column matrix, that we can interpret as an n-vector.

Then we have defined a function

$$F(v) = Av$$

that satisfies the following two important properties that characterize *linear maps*:

- $F(v + w) = F(v) + F(w)$
- $F(\alpha v) = \alpha F(w)$

Now imagine that there is another linear transformation G with associated matrix B and we want to do the composition of F with G, i.e. apply the transformation G after F has been applied:

$$G \circ F(v) = G(F(v)) = G(Av) = B(Av) = (BA)v$$

As we can see, this is also a linear map, with associated matrix BA.

2.9.1 Remark

Observe that if we apply G after F, that the respective associated matrices get multiplied in the reverse order, i.e. B goes first, then A.

2.9.2 Remark

At this very moment we can interpret matrices both as data and as data transformations.

2.10 Null space

Given an $n \times m$ matrix A, we can define yet another interesting vector space, exploiting the above described interpretation of A as a linear transformation.

The *null space* of A, denoted $N(A)$, is the vector space of all vectors that are transformed to the zero-vector by means of A. In other words:

$$N(A) = \{w \in \mathbb{R}^m \mid Aw = \vec{\mathbf{0}}\}.$$

Can you tell why $N(A)$ is a vector space at all?

2.11 TLDR;

Matrices can be thought of the mathematical abstraction of tabular data. The rank of a matrix provides us a hint of the amount of information, diversity or degrees of freedom that the matrix conveys. We can sum, scale and multiply matrices and we can think of several equivalent ways of conducting matrix multiplication. Why is matrix multiplication the way it is? We will soon learn that it has a lot to do with the fact that matrices can also be used, not only to represent data, but also to represent a type of data transformation known as linear transformations. Matrix multiplication must have been one of the most exciting and far reaching ideas in the history of mathematics.