Machine Learning and Data Analytics for Manufacturing

Linear algebra is a fundamental concept in modern mathematics. Linear algebra can be used to model many systems in nature and helps efficiently compute such models. A good understanding of linear algebra is essential for developing a fast and optimized machine learning algorithm. This document serves as a quick refresher for *Linear Algebra*. If you have little to no exposure to linear algebra, we would recommend you to consult other resources like *The Matrix Cookbook* by Peterson and Pedersen (**Free PDF Book**) or some undergrad textbooks like *Linear Algebra Done Right*. If you are already familiar with linear algebra, feel free to skip the document.

Basic Concepts and Notations

Mathematical objects:

Scalar

It's just a single number. In general, we denote them by lower-case variables in italic.

Vectors

Vector is an array of numbers. The numbers are arranged in order of the system it defines. Let, x denote an n-*dimensional* column vector with components $(x_1, x_2, ..., x_n)^T$. The vector with all the elements equal to 1 is written as $\mathbf{1}$. The ith element of the vector x is denoted by x_i. This helps in identifying a subset from original space, e.g. $S = 1, 2, n$ denotes the 1st, 2nd and nth element of the vector x.

Matrices

A matrix is a collection of scalar values arranged in a 2D array of $m \times n$ shape. A vector can be considered as a $n \times 1$ matrix. The i'th, j'th element of the matrix A can be written as A_{ij}.

Tensors

When an array has more than two axes is known as a tensor. We can identify the elements of A at co-ordinates (i, j, k) by writing A_{ijk}.

Say, we have the following systems of equation-

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ 2x_1 + 3x_2 + 4x_3 &= 5 \\ 5x_1 + 4x_2 + 3x_3 &= 12 \end{aligned} \tag{1}$$

Equation 1 can be represented compactly as-

$$Ax = b \tag{2}$$

where,

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 4 & 3 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 5 \\ 12 \end{bmatrix}$$

Using the above set of equations we can identify the following notations-

1. $A \in \mathbb{R}^{m \times n}$ denotes the entries of the A matrix which are real numbers and m denotes the rows and n denotes the columns. Here, $m = 3$ and $n = 3$.

2. $b \in \mathbb{R}^{n \times 1}$ denotes the entries of the b vector which are real numbers.

3. The *i(=2)*th element of the vector x is 5.

4. A_{23}th element of the matrix A is 4.

Identity Matrix and Diagonal Matrix

The *identity matrix* can be denoted as $I_n \in \mathbb{R}^{n \times n}$ is a square matrix which contains *ones* as the diagonal elements and otherwise *zeros*.i.e.

$$I_{ij} = \begin{cases} 1, & \text{if } i = j. \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The *diagonal matrix* is a matrix where all non-diagonal elements are zero. i.e.

$$X_{ij} = \begin{cases} x_i, & \text{if } i = j. \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Symmetric Matrix

A square matrix $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$. It is anti-symmetric if $A = -A^T$. It is easy to show that for any matrix $A \in \mathbb{R}^{n \times n}$, the matrix $A = A^T$ is symmetric and the matrix $A - A^T$ is anti-symmetric. From this it follows that any square matrix $A \in \mathbb{R}^{n \times n}$ can be represented as a sum of a symmetric matrix and an anti-symmetric matrix i.e.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) \tag{5}$$

Norms

We use *norm* to measure the size of a vector (e.g. $||x||$). In general it is denoted as $||x||_p$ and to represent L_p norm we write,

$$||x||_p = (\sum_i |x_i|^p)^{\frac{1}{p}} \tag{6}$$

for $p \in \mathbb{R}, p \geq 1$

This can be intuitively interpreted as the distance from the origin to the point x. The L_2 norm is known as the **Euclidean norm**. It's the Eucledian distance from the origin to the point of interest (say, x). A norm satisfies the following properties-

- $f(x) = 0 \rightarrow x = 0$
- $f(x+y) \leq f(x) + f(y)$ (Triangle Inequality)
- $\forall \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x)$

Matrix Operations and Manipulations

Transpose of a Matrix

The transpose of a matrix A is denoted as A^T. By performing transpose operation a matrix of $A \in \mathbb{R}^{m \times n}$ transforms into $A^T \in \mathbb{R}^{n \times m}$. Example-

say,

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 4 & 3 \end{bmatrix}$$

then,

$$A^T = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}$$

Some properties of it is, $(B^T)^T = B$, $(AB)^T = B^T A^T$ and $(ABC)^T = C^T B^T A^T$

Trace

The *trace* of a square matrix $A \in \mathbb{R}^{n \times n}$, is denoted by $tr(A)$ is the sum of the diagonal elements of the matrix A:

$$trA = \sum_{i=1}^{n} A_{ii} \tag{7}$$

Some properties of trace is-

- $A \in \mathbb{R}^{n \times n}$, $tr(A) = tr(A^T)$
- For $A, B \in \mathbb{R}^{n \times n}$, $tr(A+B) = tr(A) + tr(B)$
- $A \in \mathbb{R}^{n \times n}$, $tr(xA) = xtr(A)$
- For A, B such that AB is square, $tr(AB) = tr(BA)$.
- For A, B, C such that ABC is square, $tr(ABC) = tr(BCA) = tr(CAB)$, and so on for the product of more matrices.

Matrix Multiplication

The product of of two matrices $A \in \mathbb{R}^{l \times n}$ and $B \in \mathbb{R}^{n \times m}$ is the matrix $AB \in \mathbb{R}^{l \times m}$. This can be represented as,

$$[AB]_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk} \tag{8}$$

where, $i = 1, 2, ..., l$ and $k = 1, 2, ..., m$.

It is to be noted, in general $AB \neq BA$. When both A and B are *commute*, only then $AB = BA$ is true. A few cases of matrix multiplication can be-

Vector-Vector: Given, two distinct vector $A, B \in \mathbb{R}^m$ (The vectors are of same size), the dot product of the two matrix can be written as $x^T y$. Thus,

$$x^T y = \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Inner product are special cases of matrix multiplication. Note that it is always the case that $x^T y = y^T x$.

The **Outer product** of the matrix can be calculated as xy^T. Say, the matrices are as follows $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. Thus the resulted matrix will be $xy^T \in \mathbb{R}^{m \times n}$.

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & ... & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & ... & x_1 y_n \\ x_2 y_1 & x_2 y_2 & ... & x_2 y_n \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ x_m y_1 & x_m y_2 & ... & x_m y_n \end{bmatrix}$$

Matrix-Matrix Matrix-Matrix multiplication can be considered as a set of vector-vector products. So, to give a brief symbolic representation it will look as follows-

$$XY = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ x_{m1} & x_{m2} & ... & x_{mn} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ . & . & \cdots & . \\ . & & & \\ . & . & \cdots & . \\ . & . & \cdots & . \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} = \begin{bmatrix} x_1^T y_1 & x_1^T y_2 & ... & x_1^T y_p \\ x_2^T y_1 & x_2^T y_2 & ... & x_2^T y_p \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ x_m^T y_1 & x_m^T y_2 & ... & x_m^T y_p \end{bmatrix}$$

i.e. inner product of the ith row of X and the jth column of Y.

A few basic properties of matrix multiplication at a higher level:

- Matrix multiplication is associative: $(AB)C = A(BC)$.
- Matrix multiplication is distributive:$A(B + C) = AB + AC$.
- Matrix multiplication is, in general,not commutative; that is, it can be the case that, $AB \neq BA$.

Matrix-Vector Say, we have a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^n$, their product is a vector $y = Ax \in \mathbb{R}^m$.

$$y = Ax = \begin{bmatrix} - & a_1 & - \\ - & a_2 & - \\ - & . & - \\ - & . & - \\ - & . & - \\ - & a_m & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ . \\ . \\ . \\ a_m^T x \end{bmatrix}$$

If A is in column format then-

$$y = Ax = \begin{bmatrix} | & | & ... & | \\ a^1 & a^2 & ... & a^m \\ | & | & ... & | \end{bmatrix} x = [a^1]x + [a^2]x + ... + [a^m]x$$

Thus, y is a linear combination of the columnsof A, where the coefficients of the linear combination are given by the entries of x.

The Inverse of a Matrix

For a square matrix A is denoted as A^{-1}, and its inverse satisfies

$$A^{-1}A = I = AA^{-1} \tag{9}$$

Not all square matrices have the inverse property such that $A^{-1}A \neq I$, in which case A is *singular*. In order for a matrix to have an inverse, A must be *full rank*. Geometrically, singular matrices correspond to projections: if we transform each of the vertices v of a binary hypercube using Av, the volume of the transformed hypercube is zero. Hence, if $det(A) = 0$, the matrix A is a form of projection or 'collapse' which means A is singular. Given, a vector y and a singular transformation, A cannot be uniquely identify a vector x for which $y = Ax$. Provided the inverse exist

$$(AB)^{-1} = B^{-1}A^{-1} \tag{10}$$

for a non square matrix A such that AA^T in invertible, then the right pseudo inverse, defined as

$$A^\dagger = A^T(AA^T)^{-1} \tag{11}$$

satisfies $AA^\dagger = I$. The left pseudo inverse is given by

$$A^\dagger = (A^TA)^{-1}A^T \tag{12}$$

and it satisfies $A^\dagger A = I$.

Now, For a 2×2 matrix, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse matrix will be

$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = A^{-1} \tag{13}$$

$(ad - bc)$ is basically the determinant of the matrix A here.

References

1. Linear Algebra Review and Reference (CS229: Stanford University)
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
3. Barber, D. (2012). Bayesian reasoning and machine learning. Cambridge University Press.