Linear Algebra Summary

Zhe Li

Linear algebra plays a significant role in understanding machine learning and deep learning. Here, we summarize the main concepts of linear algebra from three different matrix decomposition perspectives[1]. Overall, the main theme is to **simplify complex matrices into simpler components**, depending on the application.

1 Solving $A\mathbf{x} = \mathbf{b}$

In order to solve $A\mathbf{x} = \mathbf{b}$, we use Gaussian Elimination (row reduction, a sequence of row-wise operations performed on the corresponding coefficients matrix), which gives the first matrix decomposition

$$A = LU \tag{1}$$

If matrix A needs row exchange so that Gaussian Elimination can be conducted, the more general LU decomposition can be written as:

$$PA = LU \tag{2}$$

where P is permutation matrix

1.1 Matrix Multiplication

How to understand matrix multiplication? take matrix multiplication $AB = C$ as example, where A is $m * n$ and B is $n * k$.

- Element-wise.
- The i^{th} column of matrix C is the linear combination of columns of matrix A with coefficient i^{th} column of matrix B.
- The i^{th} row of matrix C is the linear combination of rows of matrix B with coefficient i^{th} row of matrix A.
- Summation of i^{th} column of matrix $A * i^{th}$ row of matrix B.
- Block-wise.

[1]We largely follow MIT Linear Algebra Course.

1.2 What is $A\mathbf{x} = \mathbf{b}$?

$A\mathbf{x}$ is linear combination of columns of matrix A, which is the column space of matrix A, denoted as $C(A)$. If the vector $\mathbf{b}$ is in that column space, then to solve $A\mathbf{x} = \mathbf{b}$ is to find $\mathbf{x}$ such that the linear combination of columns of matrix A with $\mathbf{x}$ as coefficient producing the vector $\mathbf{b}$. Following this direction, we can easily summarize when there is/are solutions and how many solutions for $A\mathbf{x} = \mathbf{b}$. And how to construct the solution, we can check the following subsection the four different fundamental spaces.

1.3 The four different fundamental spaces

Corresponding to matrix A, there are four fundamental spaces:

- Columns Space, denoted as $C(A)$, a linear combination of columns of matrix A, another representation as $A\mathbf{x}$.
- Null Space, denoted as $N(A)$, which consists all solutions $\mathbf{x}$ for $A\mathbf{x} = 0$.
- Row Spaces, denoted as $C(A^T)$, a linear combination of rows of matrix A.
- Left Null Space, denoted as $N(A^T)$, which consists all solutions x for $A^T\mathbf{x} = 0$, another representation as $A^T\mathbf{x}$.

Associated with these four fundamental spaces are ranks, bases, and dimensionality.

2 How to solve $A\mathbf{x} = \mathbf{b}$ if there is no solution?

if $A\mathbf{x} = \mathbf{b}$ there is no solution, using the above terminology, the vector $\mathbf{b}$ is not in the column space of matrix A, in this case, how to get the best approximate solution?

2.1 Projection and Projection matrix

We can project the vector $\mathbf{b}$ onto the column space of matrix A. Take two vector $\mathbf{a}, \mathbf{b}$ as example to show how projection works mathematically, we can project the vector $\mathbf{b}$ onto the vector $\mathbf{a}$. Let the projection be $\mathbf{p} = x\mathbf{a}$, then error $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - x\mathbf{a}$, by using the fact that the error vector $\mathbf{e}$ perpendicular to the vector $\mathbf{a}$, that is, $\mathbf{a}^T(\mathbf{b} - \mathbf{a}x) = 0$, thus we have the projection

$$\mathbf{p} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{b} \tag{3}$$

and the projection matrix

$$\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}} \tag{4}$$

Extending the above projection to multiple vectors or column space of matrix A, we can have the projection matrix:

$$\mathbf{P} = A(A^T A)^{-1} A^T \tag{5}$$

and the projection is:

$$\mathbf{p} = A(A^T A)^{-1} A^T \mathbf{b} \tag{6}$$

Note:

- Matrix A is $m * n$ size, if A is square $n * n$ full rank matrix, we can use inverse to simplify the projection $\mathbf{P} = A(A^T A)^{-1} A^T = AA^{-1} A^T{-1} A^T = I$, identity matrix, actually, when we project the vector to a entire space ($n * n$ full rank matrix), the project matrix is indeed identity matrix.
- Why does $A^T A$ for sure have the inverse $(A^T A)^{-1}$? **If matrix A has n independent columns, the matrix $A^T A$ is non-singular and has inverse, usually in machine learning, we can craft matrix A with n independent columns by removing duplicated or redundant columns**.
- Since the final projection is in the column space of A, the projection matrix should be in this form $\mathbf{P} = A(A^T A)^{-1} A^T$, where A comes first. Just for quick check purpose.
- The projection matrix has this property, projection multiple times does not change the result, $\mathbf{P}^2 = \mathbf{P}$.

2.2 The least square

Solving $A\mathbf{x} = \mathbf{b}$ or another form $\min ||A\mathbf{x} - \mathbf{b}||^2$, by using the above projection idea, we have the well-known normal equation:

$$A^T A \hat{\mathbf{x}} = A^T \mathbf{b} \tag{7}$$

then $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$

2.3 Matrix QR decomposition

Here comes the second matrix decomposition $A = QR$, where Q is an orthonormal matrix and R is an upper triangular matrix. For the projection matrix $\mathbf{P} = A(A^T A)^{-1} A^T$, if we have a matrix Q which has same columns space with matrix A but matrix Q has orthonormal columns, that is, $Q^T Q = I$, thus we can simplify the projection matrix $\mathbf{P}$ as

$$\mathbf{P} = A(A^T A)^{-1} A^T = Q(A^T A)^{-1} A^T = QQ^T \tag{8}$$

we can use Gram-Schmidt process to get perpendicular (orthonormal) vector $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ by using projection process, thus we have

$$A = \begin{bmatrix} . & . & . & . \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ \vdots & \vdots & \vdots & \vdots \\ . & . & . & . \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & . & r_{1n} \\ \vdots & r_{22} & \vdots & r_{2n} \\ 0 & 0 & \cdots & r_{.n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & . & r_{nn} \end{bmatrix}$$

3 Eigenvalue and Eigenvector

For **square** $n*n$ matrix A, thinking matrix A representing some function/system, for some input vector $\mathbf{x}$, this function/system yields the output vector $A\mathbf{x}$ with as same direction as the input vector $\mathbf{x}$ only different scale λ. We call this λ as Eigenvalue and this input vector $\mathbf{x}$ as Eigenvector.

$$A\mathbf{x} = \lambda\mathbf{x} \tag{9}$$

Suppose that square A has n linear independent eigenvectors, put those eigenvectors into matrix S, we have

$$A = S\Lambda S^{-1} \tag{10}$$

This is the third matrix decomposition (eigen-decomposition, square matrix), which is useful in power of matrix, for example studying A^k.

There is one type of matrix which has very nice properties, symmetric matrix.

- the n eigenvalues are real.
- all eigenvectors are orthogonal.

Then, if matrix A is symmetric, we can written

$$A = S\Lambda S^{-1} = Q\Lambda Q^T = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T \tag{11}$$

This is the spectral theorem.

4 Singular Value Decomposition (SVD)

This is for any rectangle matrix A with size $m * n$. Here we are interested in tall skinny matrix A, that is $m > n$. We would like to find orthogonal bases U and V from column space and row space respectively, which has the following property:

$$AV = U\Sigma \tag{12}$$

that is $A\mathbf{v}_1 = \delta_1\mathbf{u}_1, A\mathbf{v}_2 = \delta_2\mathbf{u}_2, \cdots$.

The above can be described as orthogonal bases from row space $\rightarrow$ orthogonal bases from column space based on matrix A. Put it into decomposition formula,

$$A = U\Sigma V^T \tag{13}$$

Here matrix U is $m * m$, matrix Σ is $n * n$ and matrix V is $n * n$.

$$A = \begin{bmatrix} . & . & . & . \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ \vdots & \vdots & \vdots & \vdots \\ . & . & . & . \end{bmatrix} \begin{bmatrix} \delta_1 & 0 & . & 0 \\ \vdots & \delta_2 & \vdots & 0 \\ 0 & 0 & \vdots & \delta_n \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & . & 0 \end{bmatrix} \begin{bmatrix} . & . & . & . \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_1 & \mathbf{v}_2 & \vdots & \mathbf{v}_n \\ \vdots & \vdots & \vdots & \vdots \\ . & . & . & . \end{bmatrix}^T$$

When the rank of matrix A is r, then the above column space only has r bases, then matrix U will be $m * r$, matrix Σ will be $r * r$ and matrix V will be $n * r$.

$$A = \begin{bmatrix} . & . & . & . \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \\ \vdots & \vdots & \vdots & \vdots \\ . & . & . & . \end{bmatrix} \begin{bmatrix} \delta_1 & 0 & . & 0 \\ \vdots & \delta_2 & \vdots & 0 \\ 0 & 0 & \vdots & \delta_r \end{bmatrix} \begin{bmatrix} . & . & . & . \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_1 & \mathbf{v}_2 & \vdots & \mathbf{v}_r \\ \vdots & \vdots & \vdots & \vdots \\ . & . & . & . \end{bmatrix}^T$$

Note that matrix $A^T A$ is **square, symmetric, pos-definite** matrix, thus by using the above eigenvalue and eigenvector decomposition we have

$$A^T A = V\Sigma V^T \tag{14}$$

$$AA^T = U\Sigma U^T \tag{15}$$

By the above equation, we can compute orthonormal matrix U and V.