Linear Algebra Review

MATH 271.1: Statistical Methods

Jakov Ivan S. Dumbrique

Ateneo de Manila University

Matrix-Vector Multiplication, $\mathbf{Ax}$

Let $\mathbf{A}$ be an $n \times p$ matrix and $\mathbf{x}$ be a p-dimensional vector. Then we can think of the matrix-vector multiplication $\mathbf{Ax}$ as a linear combination of the columns of $\mathbf{A}$:

$$\begin{aligned}\mathbf{Ax} &= \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_p} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \\ &= x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_p\mathbf{a_p}\end{aligned}$$

The Four Fundamental Subspaces

An $n \times p$ matrix $\mathbf{A}$ can be defined by its four fundamental subspaces:

1. column space, $Col(\mathbf{A})$ [range of $\mathbf{A}$]
2. nullspace, $N(\mathbf{A})$ [kernel of $\mathbf{A}$]
3. row space, $Col(\mathbf{A^T})$ [corange of $\mathbf{A}$]
4. left nullspace, $N(\mathbf{A^T})$ [cokernel of $\mathbf{A}$]

The Four Fundamental Subspaces

An $n \times p$ matrix $\mathbf{A}$ can be defined by its four fundamental subspaces:

1. column space, $Col(\mathbf{A})$

$$Col(\mathbf{A}) = \{\mathbf{b} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b} \text{ for some } p\text{-vector } \mathbf{x}\}$$

2. nullspace, $N(\mathbf{A})$
3. row space, $Col(\mathbf{A}^\mathbf{T})$
4. left nullspace, $N(\mathbf{A}^\mathbf{T})$

The Four Fundamental Subspaces

An $n \times p$ matrix $\mathbf{A}$ can be defined by its four fundamental subspaces:

1. column space, $Col(\mathbf{A})$
2. nullspace, $N(\mathbf{A})$

$$N(\mathbf{A}) = \{\underline{\mathbf{x}} \in \mathbb{R}^p | \mathbf{A}\mathbf{x} = \underline{\mathbf{0}}\}$$

3. row space, $Col(\mathbf{A}^\mathbf{T})$
4. left nullspace, $N(\mathbf{A}^\mathbf{T})$

The Four Fundamental Subspaces

An $n \times p$ matrix $\mathbf{A}$ can be defined by its four fundamental subspaces:

1. column space, $Col(\mathbf{A})$
2. nullspace, $N(\mathbf{A})$
3. row space, $Col(\mathbf{A^T})$

$$Col(\mathbf{A^T}) = \{\mathbf{b} \in \mathbb{R}^p | \mathbf{A^T y} = \mathbf{b} \text{ for some } n\text{-vector } \mathbf{y}\}$$

4. left nullspace, $N(\mathbf{A^T})$

The Four Fundamental Subspaces

An $n \times p$ matrix $\mathbf{A}$ can be defined by its four fundamental subspaces:

1. column space, $Col(\mathbf{A})$
2. nullspace, $N(\mathbf{A})$
3. row space, $Col(\mathbf{A^T})$
4. left nullspace, $N(\mathbf{A^T})$

$$N(\mathbf{A^T}) = \{\mathbf{y} \in \mathbb{R}^n | \mathbf{A^T}\mathbf{y} = \mathbf{0}\}$$

How can we reconstruct a subspace?

set of vectors $\{\vec{a}, \vec{b}, \vec{c}\}$

$\vec{a} \neq a_1 \vec{b} + a_2 \vec{c}$

A subspace can be reconstructed by a set of linearly independent vectors that span the entire subspace. Such a set forms a basis for the subspace.

The number of vectors in a basis of a subspace is what we call the dimension of the subspace, $dim()$. The rank r of a matrix is just the dimension of its column space.

How are the 4 subspaces connected?

The four fundamental subspaces of a matrix are connected through their dimensions and orthogonality. This concept is captured by the following theorem:

Fundamental Theorem of Linear Algebra

Given an $n \times p$ matrix A,

1. The column space and row space have equal dimension r (rank). The nullspace has dimension $p - r$. The left nullspace has dimension $n - r$.
2. $Col(\mathbf{A^T}) = N(\mathbf{A})^{\perp}$ Orthogonal complements in $\mathbb{R}^p$
   $N(\mathbf{A^T}) = Col(\mathbf{A})^{\perp}$ Orthogonal complements in $\mathbb{R}^n$

$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y} = 0$

How are the 4 subspaces connected?

A

$Col(A^T)$

the vectors in these subspaces are in $\mathbb{R}^p$ $(\vec{b}, \vec{x} \in \mathbb{R}^p)$

Row Space

all $\vec{b} = A^T \vec{y}$

dim = r (rank)

$N(A)$

Nullspace

all $\vec{x}$: $A\vec{x} = \vec{0}$

dim = p − r

$$
\begin{aligned}
\langle \vec{x}, \vec{b} \rangle &= \vec{x}^T \vec{b} \\
&= \vec{x}^T \vec{A}^T \vec{y} \\
&= (A\vec{x})^T \vec{y} \\
&= \vec{0}^T \vec{y} \\
&= 0
\end{aligned}
$$

$Col(A)$

the vectors in these subspaces are in $\mathbb{R}^n$ $(\vec{b}, \vec{y} \in \mathbb{R}^n)$

Column Space

$\vec{b} = A\vec{x}$

dim = r (rank)

$N(A^T)$

Left Nullspace

all $\vec{y}$: $A^T\vec{y} = \vec{0}$

dim = n − r

$\langle \vec{y}, \vec{b} \rangle = 0$

Matrix-Matrix Multiplication, $\mathbf{AB}$

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices such that the matrix multiplication $\mathbf{AB}$ is defined:

$$\mathbf{AB} = \mathbf{C}$$

Matrix-Matrix Multiplication, $\mathbf{AB}$

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices such that the matrix multiplication $\mathbf{AB}$ is defined:

$$\mathbf{AB} = \mathbf{C}$$

If we flip this equation, then the multiplication turns into a factorization:

$$\underline{\mathbf{C}} = \underline{\mathbf{AB}}$$

The matrix $\mathbf{C}$ is factored into two matrices, $\mathbf{A}$ and $\mathbf{B}$. These matrices have inside information about the matrix $\mathbf{C}$. That information is not visible until you factor.

Six Important Factorizations

constrained by some conditions

Spectral decomposition

special type of #1

1. Eigenvalue decomposition $\mathbf{A} = \mathbf{P\Lambda P^{-1}}$
2. Orthogonal diagonalization $\mathbf{A} = \mathbf{Q\Lambda Q^{T}}$
3. Cholesky factorization $\mathbf{A} = \mathbf{LL^{T}}$
4. LU factorization $\mathbf{A} = \mathbf{PLU}$
5. QR factorization $\mathbf{A} = \mathbf{QR}$
6. Singular Value Decomposition $\mathbf{A} = \mathbf{U\Sigma V^{T}}$

(SVD)

SVD is at the heart of PCA ← next ML method

Eigenvalue decomposition

If $\mathbf{A}$ is a diagonalizable square matrix, then $\mathbf{A}$ can be expressed as

$\vec{x}$: eigenvector of A

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1},$$

$A\vec{x} = \lambda\vec{x}$

corresponding eigenvector of A

eigenvalue of A

where

$\mathbf{P}$ is an $n \times n$ matrix whose columns consist of n linearly independent eigenvectors of $\mathbf{A}$, and

$\mathbf{\Lambda}$ is a diagonal matrix with the corresponding eigenvalues of $\mathbf{A}$ on its main diagonal. $\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{bmatrix}$

The order of the eigenvectors used to form $\mathbf{P}$ will determine the order in which the eigenvalues appear on the main diagonal of $\mathbf{\Lambda}$.

scaling

$A\vec{x} = \lambda\vec{x}$

has the same direction as $\vec{x}$

the length is just scaled (by λ)

Eigenvalue decomposition

An example:

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

$$\begin{array}{cccc} & P & \Lambda & P^{-1} \\ \begin{bmatrix} 3 & 2 \\ 6 & -1 \end{bmatrix} = & \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} & \begin{bmatrix} -3 & 0 \\ 0 & 5 \end{bmatrix} & \begin{bmatrix} 1/4 & -1/4 \\ 3/4 & 1/4 \end{bmatrix} \end{array}$$

eigenvector of A corresponding to $\lambda_1 = -3$

eigenvector of A corresponding to $\lambda_2 = 5$

eigenvector $\lambda_1 = -3$

$\lambda_2 = 5$

$A\vec{x} = \lambda\vec{x}$

Orthogonal diagonalization

$\mathbf{A}$ is orthogonally diagonalizable if and only if $\mathbf{A}$ is symmetric:

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathbf{T}},$$

$$A = P \Lambda P^{-1}$$

where $\mathbf{Q}$ is orthogonal,
$\boldsymbol{\Lambda}$ is diagonal.

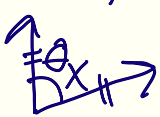
$\rightarrow$ if Q is square, $Q^{-1} = Q^T \Rightarrow Q^T Q = I$ ✓
if Q is rectangular, $Q^T Q = I$

This is just the eigenvalue decomposition of a symmetric matrix. A
As a consequence, all the eigenvalues of a symmetric matrix are real. Furthermore, all its eigenvectors are orthogonal.

$Q\vec{x}$ → preserves the length of $\vec{x}$, but it rotates/reflects $\vec{x}$

$$\|Q\vec{x}\| = \sqrt{\langle Q\vec{x}, Q\vec{x}\rangle} = \sqrt{(Q\vec{x})^T Q\vec{x}} = \sqrt{\vec{x}^T \underbrace{Q^T Q}_{I} \vec{x}} = \sqrt{\vec{x}^T \vec{x}} = \|\vec{x}\|$$

$Q\vec{x}$



Orthogonal diagonalization

An example:

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathbf{T}}$$

$$\underbrace{\begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}} = \begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}^T$$

eigenvector for $\lambda_1 = -3$

eigenvector for $\lambda_2 = 2$

$Q^{-1} = Q^T$ (finding inverse is easy!)

Cholesky factorization

Every positive definite matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$$

where $\mathbf{L}$ is lower triangular with positive diagonal elements.

Cholesky factorization

Every positive definite matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$$

where $\mathbf{L}$ is lower triangular with positive diagonal elements.

Cholesky factorization is useful in solving linear equations, linear least-squares and least-norm problems.

Cholesky factorization

An example:

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$$

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

LU factorization

Every square nonsingular matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{PLU},$$

where
$\mathbf{P}$ is an $n \times n$ permutation matrix
$\mathbf{L}$ is a unit lower traingular $n \times n$ matrix
$\mathbf{U}$ is a nonsingular upper triangular $n \times n$ matrix

LU factorization is useful in solving linear equations.

LU factorization

An example:

$$\mathbf{A} = \mathbf{PLU}$$

$$\begin{bmatrix} 0 & 5 & 5 \\ 2 & 9 & 0 \\ 6 & 8 & 8 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & -19/5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 9 & 0 \\ 0 & 5 & 5 \\ 0 & 0 & 27 \end{bmatrix}$$

QR factorization

If $\mathbf{A}$ is an $n \times p$ matrix with a zero nullspace, then it can be factored as

$$\mathbf{A} = \mathbf{QR},$$

where
$\mathbf{Q}$ is an $n \times p$ orthogonal matrix, and
$\mathbf{R}$ is a $p \times p$ upper triangular matrix with positive diagonal elements

QR factorization

If $\mathbf{A}$ is an $n \times p$ matrix with a zero nullspace, then it can be factored as

$$\mathbf{A} = \mathbf{Q}\mathbf{R},$$

where
$\mathbf{Q}$ is an $n \times p$ orthogonal matrix, and
$\mathbf{R}$ is a $p \times p$ upper triangular matrix with positive diagonal elements

QR factorization is useful in solving linear least-squares and least-norm problems.

QR factorization

An example:

$$\mathbf{A} = \mathbf{QR}$$

$$\frac{1}{5}\begin{bmatrix} 3 & -6 & 26 \\ 4 & -8 & -7 \\ 0 & 4 & 4 \\ 0 & -3 & -3 \end{bmatrix} = \begin{bmatrix} 3/5 & 0 & 4/5 \\ 4/5 & 0 & -3/5 \\ 0 & 4/5 & 0 \\ 0 & -3/5 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 5 \end{bmatrix}$$

Singular Value Decomposition

Given any $n \times p$ matrix $\mathbf{A}$, its singular value decomposition is given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$$

where

$\mathbf{U}$ is orthonormal,
$\mathbf{\Sigma}$ is diagonal, and
$\mathbf{V}$ is orthonormal.

Singular Value Decomposition

Given any $n \times p$ matrix $\mathbf{A}$, its singular value decomposition is given by

$$\mathbf{A} = \mathbf{U\Sigma V^T}$$

where

$\mathbf{U}$ is orthonormal,
$\mathbf{\Sigma}$ is diagonal, and
$\mathbf{V}$ is orthonormal.

SVD is useful in a lot of applications. In this class, we will focus on its application on *principal components analysis*.

Singular Value Decomposition

An example:

$$\mathbf{A} = \mathbf{U\Sigma V^T}$$

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$