Lecture 2

A Short Review of Matrix Algebra

Reading: Zelterman, 2015 Chapter 4; Izenman, 2008 Chapter 3.1-3.2

DSA 8070 Multivariate Analysis

Whitney Huang
Clemson University

Notes

Agenda

1. Motivation
2. Basic Matrix Concepts
3. Some Useful Matrix Tools/Facts

Notes

Why Matrix Algebra?

Data:

	crim	zn	indus	chas	nox	rm
1	0.00632	18	2.31	0	0.538	6.575
2	0.02731	0	7.07	0	0.469	6.421
3	0.02729	0	7.07	0	0.469	7.185
4	0.03237	0	2.18	0	0.458	6.998
5	0.06905	0	2.18	0	0.458	7.147
6	0.02985	0	2.18	0	0.458	6.430

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Summary Statistics:

$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{bmatrix} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1}$ is the sample mean vector,

and $\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \cdots & \cdots & \cdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} = \frac{1}{n-1}\boldsymbol{X}^T(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\boldsymbol{X}$ is the sample covariance matrix

$\Rightarrow$ Many matrix algebra techniques will be applied to this matrix in multivariate analysis

Notes

Vectors

- A column array of p elements is called a vector of dimension p and is written as

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

  $\Rightarrow$ Each observation in a multivariate dataset is a p-dimensional vector (e.g., exam scores in math, science, and writing).

- The transpose of the column vector $\mathbf{a}$ is a row vector

$$\boldsymbol{a}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}$$

- $L_{\mathbf{a}}^{-1}\mathbf{a} = \frac{\mathbf{a}}{\sqrt{\sum_{i=1}^n a_i^2}}$ is called a unit vector

Notes

Vectors in Multivariate Analysis

- **Column vector (observation):** Each observation $\mathbf{x}_i \in \mathbb{R}^p$ is a $p \times 1$ column; stacking rows yields the data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$

- **Transpose:** Enables matrix operations such as inner products and summary statistics, e.g., $\mathbf{x}^\top \mathbf{y}$ (inner product), $\mathbf{a}^T\mathbf{x}$ (linear combination), $\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X}^T\mathbf{1}$ (mean), $\mathbf{X}^T\mathbf{X}$ (cross-product for covariance)

- **Unit vector:** normalize $\mathbf{x}$ to $\mathbf{x}/\|\mathbf{x}\|$ (length 1) to remove scale and compare directions

A Short Review of Matrix Algebra
CLEMSON UNIVERSITY
Motivation
Basic Matrix Concepts
Some Useful Matrix Tools/Facts

Notes

Matrices

- A matrix A is an array of elements a_{ij} with n rows and p columns:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

- The transpose $\boldsymbol{A}^T$ has p rows and n columns. The j-th row of $\boldsymbol{A}^T$ is the j-th column of $\boldsymbol{A}$

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}$$

- Key matrices in multivariate analysis: data matrix $\boldsymbol{X}$, covariance/correlation $\boldsymbol{S}, \boldsymbol{R}$, and eigen decomposition

A Short Review of Matrix Algebra
CLEMSON UNIVERSITY
Motivation
Basic Matrix Concepts
Some Useful Matrix Tools/Facts

Notes

Covariance Matrices

- Covariance Matrix

$$\mathbf{\Sigma} = \mathbb{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] = \underbrace{\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}}_{\textbf{population}\text{ covariance matrix}}, \quad \boldsymbol{S} = \frac{1}{n-1}\left(\boldsymbol{X}-\mathbf{1}\,\bar{\mathbf{x}}^T\right)^T\left(\mathbf{X}-\mathbf{1}\,\bar{\mathbf{x}}^T\right) = \underbrace{\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \cdots & \cdots & \cdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}}_{\textbf{sample}\text{ covariance matrix}}$$

- Since $\sigma_{jk} = \sigma_{kj}$ (likewise $s_{jk} = s_{kj}$) for all $j \neq k \Rightarrow \mathbf{\Sigma}$ and $\boldsymbol{S}$ are symmetric
- $\mathbf{\Sigma}$ and $\boldsymbol{S}$ are also non-negative definite $\Rightarrow$ Any linear combination of the variables has nonnegative variance

Notes

Identity Matrix and Inverse Matrix

A Short Review of Matrix Algebra
CLEMSON UNIVERSITY
Motivation
Basic Matrix Concepts
Some Useful Matrix Tools/Facts

- An identity matrix, denoted by $\mathbf{I}$, is a square matrix with 1's along the diagonal and 0's everywhere else. For example,

$$\mathbf{I}_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Consider two square matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension. If

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I},$$

  then $\mathbf{B}$ is the inverse of $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$.

- The inverse matrix is used in multivariate analysis for standardization (e.g., Mahalanobis distance).

Notes

Orthogonal Matrices

A Short Review of Matrix Algebra
CLEMSON UNIVERSITY
Motivation
Basic Matrix Concepts
Some Useful Matrix Tools/Facts

- A square matrix $\mathbf{Q}$ is orthogonal if

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = I$$

- If $\mathbf{Q}$ is orthogonal, its rows and columns have unit length (i.e., $L_{\mathbf{q}_j} = 1$) and are mutually perpendicular (i.e., $\mathbf{q}_j^T\mathbf{q}_k = 0$ for any $j \neq k$)
- **Example:**

$$\mathbf{Q} = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}$$

- Orthogonal matrices are used in multivariate analysis for rotations and transformations

Notes

Eigenvalues and Eigenvectors

- A square matrix A has an eigenvalue λ with corresponding eigenvector $\mathbf{x} \neq 0$ if

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

 The eigenvalues of $\mathbf{A}$ are the solution to $|\mathbf{A} - \lambda I| = 0$

- A normalized eigenvector is denoted by $\mathbf{e}$ with $\mathbf{e}^T\mathbf{e} = 1$

- A $p \times p$ matrix A has p pairs of eigenvalues and eigenvectors

$$\lambda_1, \mathbf{e}_1 \quad \lambda_2, \mathbf{e}_2 \quad \cdots \quad \lambda_p, \mathbf{e}_p$$

Notes

Spectral Decomposition

- Eigenvalues and eigenvectors will play an important role in DSA 8070. For example, principal components are based on the eigenvalues and eigenvectors of sample covariance matrices

- The spectral decomposition of a $p \times p$ symmetric matrix $\mathbf{A}$ is $\mathbf{A} = \lambda_1\mathbf{e}_1\mathbf{e}_1^T + \lambda_2\mathbf{e}_2\mathbf{e}_2^T + \cdots + \lambda_p\mathbf{e}_p\mathbf{e}_p^T$. Matrix form:

$$\underbrace{\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p\end{bmatrix}}_{P} \underbrace{\begin{bmatrix}\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p\end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p\end{bmatrix}^T}_{P^T}$$

- In PCA, let $\mathbf{A}$ be the covariance matrix; sort $\lambda_1 \geq \cdots \geq \lambda_p$: eigenvectors $\mathbf{e}_j \Rightarrow$ principal components, eigenvalues $\lambda_j \Rightarrow$ variances

Notes

Determinant, Trace, and Rank

- The trace of a $p \times p$ matrix $\mathbf{A}$ is the sum of its diagonal elements, i.e., $\text{trace}(\mathbf{A}) = \sum_{i=1}^{p} a_{ii}$.

- The trace of a square, symmetric matrix $\mathbf{A}$ is the sum of its eigenvalues, i.e., $\text{trace}(\mathbf{A}) = \sum_{i=1}^{p} a_{ii} = \sum_{i=1}^{p} \lambda_i$

- The determinant of a square, symmetric matrix A is the product of its eigenvalues, i.e., $|\mathbf{A}| = \prod_{i=1}^{p} \lambda_i$

- The rank of a matrix $\mathbf{A}$ is the dimension of the vector space spanned by its rows (or equivalently, its columns). It is equal to the number of nonzero eigenvalues of $\mathbf{A}$

Notes

Determinant, Trace, and Rank: Applications in Multivariate Analysis

- The determinant of the covariance matrix is used to measure the generalized variance of a multivariate distribution.
- The trace of the covariance matrix represents the total variance across all variables.
- The rank of a data matrix (or covariance matrix) indicates the effective dimensionality of the data, revealing linear dependence among variables

Notes

Positive Definite Matrix

- For a $p \times p$ symmetric matrix $\mathbf{A}$ and a vector $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_p]^T$ the quantity $\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_ix_j$ is called a quadratic form
- If $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for any vector $\mathbf{x}$, both $\mathbf{A}$ and the quadratic form are said to be non-negative definite $\Rightarrow$ all the eigenvalues of $\mathbf{A}$ are non-negative
- If $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for any vector $\mathbf{x} \neq \mathbf{0}$, both $\mathbf{A}$ and the quadratic form are said to be positive definite $\Rightarrow$ all the eigenvalues of $\mathbf{A}$ are positive
- In multivariate analysis, the covariance matrix must be positive definite to ensure valid Mahalanobis distances, PCA, and multivariate normal distributions

Notes

Distance and Quadratic Forms

- For $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_p]^T$ and a $p \times p$ positive definite matrix $\mathbf{A}$,

$$d^2 = \mathbf{x}^T\mathbf{A}\mathbf{x}$$

when $\mathbf{x} \neq 0$. Thus, a positive definite quadratic form can be interpreted as a squared distance of $\mathbf{x}$ from the origin and vice versa
- The squared distance from $\mathbf{x}$ to a fixed point $\boldsymbol{\mu}$ is given by the quadratic form

$$(\mathbf{x} - \boldsymbol{\mu})^T A(\mathbf{x} - \boldsymbol{\mu})$$

- In multivariate analysis, such quadratic forms are used to define the Mahalanobis distance, construct confidence ellipsoids, and perform discriminant analysis

Notes

Distance and Quadratic Forms (Cont'd)

- We can interpret distance in terms of eigenvalues and eigenvectors of $\mathbf{A}$. any point $\mathbf{x}$ at constant distance c from the origin satisfies

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(\sum_{j=1}^{p}\lambda_j\mathbf{e}_j\mathbf{e}_j^T)\mathbf{x} = \sum_{j=1}^{p}\lambda_j(\mathbf{x}^T\mathbf{e}_j)^2 = c$$

- Note that the point $\mathbf{x} = c\lambda_1^{-\frac{1}{2}}\mathbf{e}_1$ is at a distance c (in the direction of $\mathbf{e}_1$) from the origin because it satisfies $\mathbf{x}^T\mathbf{A}\mathbf{x} = c^2$
- The same is true for points $\mathbf{x} = c\lambda_j^{-\frac{1}{2}}\mathbf{e}_j,\ j = 2, \cdots, p.$ Thus, all points at distance c lie on an ellipsoid with axes in the directions of the eigenvectors and with lengths proportional to $\lambda_j^{-\frac{1}{2}}$

Notes

Square-Root Matrices

- Spectral decomposition of a positive definite matrix $\mathbf{A}$ yields

$$\mathbf{A} = \sum_{j=1}^{p}\lambda_j\mathbf{e}_j\mathbf{e}_j^T = \mathbf{P}\Lambda\mathbf{P}^T,$$

with $\Lambda_{p\times p} = \mathrm{diag}(\lambda_j)$, all $\lambda_j > 0$, and $\mathbf{P}_{p\times p} = \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_p\end{bmatrix}$ an orthonormal matrix of eigenvectors. Then

$$\mathbf{A}^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}^T = \sum_{j=1}^{p}\frac{1}{\lambda_j}\mathbf{e}_j\mathbf{e}_j^T$$

- With $\Lambda^{\frac{1}{2}} = \mathrm{diag}(\lambda_j^{\frac{1}{2}})$, a square-root matrix is

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{P}\Lambda^{\frac{1}{2}}\mathbf{P}^T = \sum_{j=1}^{p}\sqrt{\lambda_j}\mathbf{e}_j\mathbf{e}_j^T$$

Notes

Partitioning Random vectors

- If we partition the $p \times 1$ random vector $\boldsymbol{X}$ into two components $\boldsymbol{X}_1, \boldsymbol{X}_2$ of dimensions $q \times 1$ and $(p-q) \times 1$ respectively, then the mean vector and the variance-covariance matrix need to be partitioned accordingly
- Partitioned mean vector:

$$\mathbb{E}[\boldsymbol{X}] = \mathbb{E}\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix} = \begin{bmatrix}\mathbb{E}[\boldsymbol{X}_1]\\ \mathbb{E}[\boldsymbol{X}_2]\end{bmatrix} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$$

- Partitioned covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix}\mathrm{Var}(\boldsymbol{X}_1) & \mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)\\ \mathrm{Cov}(\boldsymbol{X}_2, \boldsymbol{X}_1) & \mathrm{Var}(\boldsymbol{X}_2)\end{bmatrix} = \begin{bmatrix}\underbrace{\boldsymbol{\Sigma}_{11}}_{q\times q} & \underbrace{\boldsymbol{\Sigma}_{12}}_{q\times(p-q)}\\ \underbrace{\boldsymbol{\Sigma}_{21}}_{(p-q)\times q} & \underbrace{\boldsymbol{\Sigma}_{22}}_{(p-q)\times(p-q)}\end{bmatrix}$$

Notes

Summary: Matrix Algebra in Multivariate Analysis

- Data as a matrix $\boldsymbol{X}$; each row is an observation, each column a variable
- Sample mean vector and covariance matrix are matrix expressions
- Eigenvalues/eigenvectors ⇒ PCA, factor analysis, canonical correlation
- Quadratic forms ⇒ Mahalanobis distance, hypothesis testing

In the next lecture, we will learn:

- Multivariate Normal Distribution
- Copula Models and Non-parametric Density Methods

Notes

Notes

Notes