분석적 사고

전종준
서울시립대학교 통계학과

September 15, 2018

Abstract

키워드: 선형대수, 미적분학, 회귀분석, 탐색적 자료분석

1 자료와 집합

2 벡터와 벡터공간

숙제 제출: 자료의 차원에 대한 논의

3 행렬과 데이터

3.1 행렬을 이용한 데이터의 표현

<u>다변량 자료의 표현</u>

- 다변량 자료를 엑셀 시트에 표기를 해보시오
- 관습적인 테이블 데이터의 표현
- 첨자를 사용한 데이터의 표현

<u>행렬 쓰기</u>

- 행렬과 원소 (벡터 만들기를 참고)
- 데이터 테이블 내에서 행과 열의 의미
- $\mathbf{X}$: $n \times p$ matrix of which the ith and the jth elements is given by x_{ij}.

  (여기서 특별히 원소가 $\mathbb{R}$이고 $n \times p$ matrix를 모두 모아놓은 집합을 $\mathcal{M}_{n\times p}$ 라고 표기하고 $(\mathbf{X})_{ij}$를 행렬 $\mathbf{X}$의 i행 j열 원소라고 표시하자.)

3.2 특별한 행렬

- rectangular matrix
- identity matrix
- diagonal matrix (diagonal elements)

definition: upper triangular matrix, sparse matrix.

3.3 행렬의 계산

<u>더하기 연산</u>

- $A, B \in \mathcal{M}_{n\times p}$: $A+B$는 다음과 같이 정의한다.

<u>(스칼라) 곱의 연산</u>

- $A \in \mathcal{M}_{n\times p}$ 그리고 $c \in \mathbb{R}$ 에 대해서 $c \times A$는 다음과 같이 정의한다.

3.4 행렬의 변환

<u>행렬의 곱</u>

- $A \in \mathcal{M}_{n\times p}$ 와 $\mathbf{B} \in \mathcal{M}_{m\times n}$을 생각한다.
- $\mathbf{B}A \in \mathcal{M}_{m\times p}$ 이고 $(\mathbf{B}A)_{ij} = \sum_{k=1}^{n} (\mathbf{B})_{ik}(A)_{kj}$로 주어진다.
- 행렬의 곱은 집합 $\mathcal{M}_{n\times p}$위에서 정의된 연산이 아니라 변환으로 생각할 수 있다. 어떤 변환인가? (hist: 입력 A, 출력 BA)
- R을 이용한 변환의 특성 파악하기

$$\mathbf{B}(c_1A_1 + c_2A_2) = c_1\mathbf{B}A_1 + c_2\mathbf{B}A_2$$

<u>전치: transpose</u>

- 전치를 R을 통해 확인하기
- 전치는 어떤 변환인가?
- 전치의 표기 A^T
- 전치의 성질 확인하기

$$(BA)^T = A^T B^T$$

- 열 벡터의 표현과 계산

3.5 역행렬

- $A \in \mathcal{M}_{p\times p}$ 어떤 $B \in \mathcal{M}_{p\times p}$ 가 있어서

$$AB = BA = I$$

인 B가 있으면 B를 역행렬이라고 부르고 A^{-1}라고 표시한다.

3.6 역행렬을 이용한 방정식 풀기

<u>이원 일차 연립방정식 쓰기</u>

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ x_1 - x_2 &= 1 \end{aligned}$$

Then,

$$\begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

<u>삼원 일차 연립방정식 쓰기</u>
<u>p원 일차 연립방정식의 일반형 쓰기</u>

$$\begin{pmatrix} a_{11} & \cdots & a_{1p} \\ a_{21} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$$

<u>역행렬을 이용해서 방정식 풀기</u>

$$Ax = b \to A^{-1}Ax = A^{-1}b$$

<u>R로 계산해보기</u>

- solve function

연습/숙제 $p = 4$ 개의 변량을 가진 자료를 $n = 5$개 가지고 있다. 예를 들어 한 명의 사람에 대해서 나이, 성별, 키, 몸무게 측정하였고 총 5명의 자료를 모았다고 가정하자. 첫번째 사람의 자료를

$$\mathbf{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \end{pmatrix}$$

이라고 표시한다. (관습적으로 벡터는 열벡터로 다룸). 여기서 $x_{11}, x_{12}, x_{13}, x_{14}$는 첫번째 사람의 나이, 성별, 키, 몸무게다. k번째 사람 $(k = 1, \cdots, 5)$은 다음과 같이 쓸 수 있을 것이다.

$$\mathbf{x}_k = \begin{pmatrix} x_{k1} \\ x_{k2} \\ x_{k3} \\ x_{k4} \end{pmatrix}$$

많은 경우 열벡터를 표시할 대 종이의 낭비를 줄이고자 다음과 같이 표현한다.

$$\mathbf{x}_k = (x_{k1}, x_{k2}, x_{k3}, x_{k4})^T$$

사람의 데이터를 행렬로 모아서 표현을 하면 다음과 같을 것이다.

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_{100}^T \end{pmatrix}$$

그리고 사실 이는 다음과 같은 행렬의 전치임을 알 수 있다.

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{10} \end{pmatrix}^T$$

한편 어떤 벡터 $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_4)^T$ 가 있을때, $\mathbf{X}\beta$ 를 계산해보면

$$\mathbf{X}\beta = \begin{pmatrix} \mathbf{x}_1^T\boldsymbol{\beta} \\ \mathbf{x}_2^T\boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_{100}^T\boldsymbol{\beta} \end{pmatrix}$$

임을 알 수 있다.

여기서 $\mathbf{X}$를 다시 써보자.

$$\mathbf{X} = (X_1, X_2, X_3, X_4)$$

여기서 $X_j = (x_{1j}, \cdots, x_{10j})^T$ 이다. 그리고

$$\mathbf{X}\boldsymbol{\beta} = \beta_1 X_1 + \cdots + \beta_4 X_4$$

임을 알 수 있다.

지금 까지 설명한 내용을 실제 예를 들어서 확인해보고, 행렬을 이용한 데이터 연산에 대한 본인의 의견을 적어라.