분석적 사고

전종준
서울시립대학교 통계학과

October 1, 2018

Abstract

키워드: 선형대수, 미적분학, 회귀분석, 탐색적 자료분석

1 자료와 집합

2 벡터와 벡터공간

3 행렬과 데이터

숙제 제출: 행렬을 이용한 데이터 표현과 연산에 대한 논의

4 선형변환과 이차형식

4.1 자료의 변환

행이 n이고 열이 p인 데이터 행렬 $\mathbf{X}$을 생각한다. 3장에서 데이터 행렬 $\mathbf{X}$에 대한 표현을 살펴보았다.

- $\mathbf{x}_i$가 i번째 관측치를 표현하는 열벡터인 경우: $\mathbf{x}_i \in \mathbb{R}^p$ 이며 데이터 행렬은 다음과 같이 표현된다.

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

- X_j가 j번째 변수를 표현하는 열벡터인 경우: $X_j \in \mathbb{R}^n$ 이며 데이터 행렬은 다음과 같이 표현된다.

$$\mathbf{X} = (X_1, \cdots, X_p)$$

그리고 $\mathbf{X}\boldsymbol{\beta}$에 대한 표현형도 살펴보았다.

$$\mathbf{X}\boldsymbol{\beta} =$$

- $\mathbf{x}_i^T\boldsymbol{\beta}$를 풀어서 써 보아라.

- $\mathbf{x}_i^T\mathbf{x}_i$를 풀어서 써 보아라.

- $\mathbf{x}_i$를 p차원 공간상의 한 점이라고 하면 $\sqrt{\mathbf{x}_i^T\mathbf{x}_i}$의미는 무엇인가?

- 유클리디안 거리의 축약표현: $\|\mathbf{x}_i\|_2$ (p차원 벡터 $\mathbf{x}_i$의 길이)

(question) $\mathbf{X}\boldsymbol{\beta}$는 어디에 속하는 원소인가? (어떤 형태의 값인가?)
(참고) column space: $\mathbf{X} \in \mathcal{M}_{n,p}$ 라고 하자.

$$\mathcal{C}(\mathbf{X}) = \{y \in \mathbb{R}^n : y = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^p\}$$

를 행렬 $\mathbf{X}$가 생성한 column space라고 한다.

관심이 있는 변수 y 가 있고 관심있는 변수를 설명하고자 하는 변수 $\mathbf{x}$ 가 있다고 하자. (예: 주택의 가격상승률(y), 설명변수 p개. 즉 $\mathbf{x} \in \mathbb{R}^p$) 과거의 i 시점의 주택 가격 상승률 $y_i \in \mathbb{R}$과 그에 대응되는 설명변수 자료 $\mathbf{x}_i \in \mathbb{R}^p$ n개 가있어 다음과 같이 표시하였다.

- $\mathbf{y} = (y_1, \cdots, y_n)^T$

- $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^T$

- 어떤 벡터 $\boldsymbol{\beta} \in \mathbb{R}^p$에 대해서 $(y - \mathbf{X}\boldsymbol{\beta})$ 의 의미를 설명하여라.

- $\|(y - \mathbf{X}\boldsymbol{\beta})\|_2^2 = (y - \mathbf{X}\boldsymbol{\beta})^T(y - \mathbf{X}\boldsymbol{\beta})$ 임을 확인하고 그 의미를 설명하여라.

- 그래프를 통한 확인의 연습

<u>$\mathbf{X}\boldsymbol{\beta}$에 대한 고찰</u>
$\mathbf{X} \in \mathcal{M}_{n,p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ $(p < n)$ 이라고 하자.

- $\mathbf{X}$를 고정하고 $\boldsymbol{\beta}$를 어떤 입력값으로 생각해보자. $\boldsymbol{\beta}$는 어떤 공간에서의 입력값인가?

- $\mathbf{X}\boldsymbol{\beta}$를 출력값으로 생각해보자. 이 값은 어떤 공간의 원소인가?

- $\mathbf{X}$를 어떤 함수라고 생각하면 어디에 있는 입력값을 어디에 있는 출력값으로 보내는가? (선형필터, 선형변환)

- $\mathbf{X}\boldsymbol{\beta}$를 $\boldsymbol{\beta}$에 대한 함수로 생각하자. 이 함수는 β_j의 제곱항 또는 $\beta_j\beta_k$ $(j \neq k)$항을 을 포함하는가?

4.2 다차원 대응

$f : \mathbb{R}^p \mapsto \mathbb{R}^q$ 인 대응을 생각해보자. 먼저 f라는 함수 혹은 대응이 함수가 되기 위한 조건은 무엇인가? 함수의 정의를 통해서 확인해보자.

f의 이미지를 다음과 같이 나누어서 생각해볼 수 있다.

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_q(\mathbf{x}) \end{pmatrix}$$

여기서 $f_j : \mathbb{R}^p \mapsto \mathbb{R}$ 이다. 이과 같이 이미지가 다차원 벡터인 경우에는 각 원소를 대응시키는 함수들을 살펴봄으로써 함수가 어떻게 이미지를 대응시키고 있는지 파악할 수 있다.

예) $f(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$라고 하자. 여기서 $\mathbf{X} \in \mathcal{M}_{q\times p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ 다. f의 이미지를 벡터형식으로 써보고, 위에서 정의한 f_j for $j = 1, \cdots, q$가 어떤 형식을 쓰어지는지 확인하여라.

4.3 이차형식

행렬의 곱에 대한 계산

$$(A + B)(C + D) = AC + AD + BC + BD$$

전치 관련 계산

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

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

R을 통해서 확인해보아라.

$(y - \mathbf{X}\boldsymbol{\beta})^T(y - \mathbf{X}\boldsymbol{\beta})$ 를 계산해보자

$$\begin{aligned} (y - \mathbf{X}\boldsymbol{\beta})^T(y - \mathbf{X}\boldsymbol{\beta}) &= (y^T - (\mathbf{X}\boldsymbol{\beta})^T)(y - \mathbf{X}\boldsymbol{\beta}) \\ &= (y^T - \boldsymbol{\beta}^T\mathbf{X}^T)(y - \mathbf{X}\boldsymbol{\beta}) \\ &= (y^T y - y^T\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T y + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

- $z \in \mathbb{R}$ 이면 $z = z^T$ 이다.
- $\boldsymbol{\beta}^T \mathbf{X}^T y \in \mathbb{R}$ 인가?
- $\boldsymbol{\beta}^T \mathbf{X}^T y = y^T \mathbf{X} \boldsymbol{\beta}$인가?
- $\boldsymbol{\beta}^T \mathbf{X}^T y = (\mathbf{X}^T y)^T \boldsymbol{\beta}$인가?

$$(y - \mathbf{X}\boldsymbol{\beta})^T (y - \mathbf{X}\boldsymbol{\beta}) = (y^T y - 2y^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta})$$

임을 확인하자.

- A가 $p \times p$ 행렬이라 하자. $\boldsymbol{\beta}^T A \boldsymbol{\beta}$의 값을 계산해보고 최고차 항을 확인해보아라.
- $b \in \mathbb{R}^p$ 일때 $b^T \boldsymbol{\beta}$의 최고차 항을 확인하여라.

숙제1: 함수의 이차형식을 조사하여라.
숙제2: 이차형식을 구체적인 예를 3개 만들어보아라. $n = 3, p = 2$ 일 때 y와 $\mathbf{X}$는 구체적인 숫자로 놓고 $\boldsymbol{\beta}$는 변수로 놓은 후, $(y - \mathbf{X}\boldsymbol{\beta})^T (y - \mathbf{X}\boldsymbol{\beta})$를 계산하여

$$(y - \mathbf{X}\boldsymbol{\beta})^T (y - \mathbf{X}\boldsymbol{\beta}) = (y^T y - 2y^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta})$$

임을 확인하여라. $\mathbf{X}^T \mathbf{X}$는 어떤 특성을 가진 행렬인가?