분석적 사고

전종준
서울시립대학교 통계학과

December 7, 2018

Abstract

키워드: 선형대수, 미적분학, 회귀분석, 탐색적 자료분석

1 자료와 집합

2 벡터와 벡터공간

3 행렬과 데이터

4 선형변환과 이차형식

5 미분과 적분

6 벡터미분

7 확률과 기대값

8 경험을 통한 학습

8.1 자료의 중심찾기

일차원 자료 X의 중심을 찾는 문제를 생각해보자. 만약 분포 F를 (혹은 pdf f를) 알고 있다면 다음과 같은 방법을 생각해볼 수 있을 것이다.

- $\mu \in \mathbb{R}$: 예상하는 중심의 위치
- $(X-\mu)^2$: 자료와 예상한 위치가 떨어진 크기
- $\mathrm{E}(X-\mu)^2 = \int (x-\mu)^2 f(x)dx$: 자료와 예상한 위치가 떨어진 크기의 평균(기대값)

우리는 아마도 $\mathrm{E}(X-\mu)^2$를 가장 최소화 하는 값을 선택할 것이다.

위 내용과 관련된 용어는 다음과 같다.

- $\mu = \delta(x) \in \mathbb{R}$ 의사결정 (Decision function)
- $l(\mu, x) = (x-\mu)^2$: 손실함수 (loss function)
- $R(\mu) = \mathrm{E}(l(\mu, X))$: 위험함수 (Risk function)

위 예에 대한 답을 살펴보자. 위험함수는 μ에 대한 함수이며 항상 양수 값을 가질 것이다. 이 함수를 μ에 대해 미분하면

$$\frac{\partial}{\partial \mu}\left(\int (x-\mu)^2 f(x)dx\right) = \int 2(\mu - x)f(x)dx$$

이고 미분값이 0 이되는 μ^*를 찾으면

$$\mu^* = \int x f(x)dx$$

가 나온다.

실제 문제에서는 분포함수가 알려져 있지 않으므로 이 μ^*는 이론적으로만 알 수 있는 값이다. 만약 어떤 분포 F 일차원 자료 X_i for $i = 1, \cdots, n$ 가 있을 때 μ를 어떻게 알아낼 수 있을까? 유한한 자료하에서 진짜 μ를 알아낼 수는 없겠지만, 다음과 같은 방법으로 μ를 근사할 수 있을 것이다.

- $R(\mu)$ 대신

$$\hat{R}(\mu) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

을 사용한다.

- $\hat{R}(\mu)$를 최소화 하는 $\hat{\mu}$를 사용한다. (계산을 통해 아래 결과가 나옴을 확인하여라)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

여기서 $\hat{R}(\mu)$를 경험적 위험함수라고 부르고 $\hat{\mu}$를 μ^*의 추정량이라고 부른다.

- $\hat{\mu}$는 정말로 μ^*의 근사인가?

$$\hat{\mu} \to \mu^*$$

as $n \to \infty$.

8.2 회귀모형

- 조건부 기대값과 예: 어떤 입력 변수 $X = (X_1, \cdots, X_p)^T$가 주어졌을 때, 반응변수 Y의 평균이 X에 의존하는 모형 (예를 들어보자). 특별히 선형모형에 대한 식을 나타내면 다음과 같다.

$$\mathrm{E}(Y|X) = g(X)$$

(여기서 $Y \in \mathbb{R}$, $X \in \mathbb{R}^p$ 인 확률변수)

조건부 기대값을 $X \in \mathbb{R}$에 의존하는 어떤 함수로 볼 수 있다.

많은 경우 다음과 (Y, X)에 대한 생성모형을 다음과 같이 정의한다.

$$Y = g(X) + \epsilon,$$

where $\epsilon \sim (0, \sigma^2)$. (여기서 생성모형이란 데이터가 생성되는 구조를 기술하는 확률모형을 말한다.)

- 의사결정 함수: $\delta : \mathbb{R}^p \mapsto \mathbb{R}$
- 손실함수 $l(y,z) = (y-z)^2$
- 위험함수 $R(\delta) = \mathrm{E}(l(Y,\delta(X)) = \mathrm{E}((Y-\delta(X)^2)$ (여기서 δ가 함수임을 유의!)

어떤 함수($\delta : \mathbb{R}^p \mapsto \mathbb{R}$) 이 위에서 정의한 위험함수를 최소화 할까?

$$
\begin{aligned}
\mathrm{E}(Y-\delta(X))^2) &= \mathrm{E}(Y-\mathrm{E}(Y|X)+\mathrm{E}(Y|X)-\delta(X))^2) \\
&= (\mathrm{E}(Y-\mathrm{E}(Y|X))^2) + \underbrace{\mathrm{E}(Y-\mathrm{E}(Y|X)(\mathrm{E}(Y|X)-\delta(X))}_{=0} \\
&\quad + \underbrace{\mathrm{E}((\mathrm{E}(Y|X)-\delta(X))^2)}_{\geq 0}
\end{aligned}
$$

$\delta(X) = \mathrm{E}(Y|X)$로 주어질 때 위험함수는 최소화 된다. 만약 $g(X) = X^T\boldsymbol{\beta}$라고 가정해보자. (p차원 parametrization)

- $\delta(X) = X^T\boldsymbol{\beta}$
- $R(\boldsymbol{\beta}) = \mathrm{E}(Y - X^T\boldsymbol{\beta})^2$

경험적 위험함수의 최소값은 다음과 같이 주어진다.

$$\boldsymbol{\beta}^* = \mathrm{E}(XX^T)^{-1}\mathrm{E}(XY).$$

<u>조건부 평균의 추정</u>

선형모형의 경우에 조건부 평균의 추정 문제를 다룬다.

$$\mathrm{E}Y|X = X^T\boldsymbol{\beta}$$

경우에 따라서는 Y의 조건부 평균값의 범위를 제한해야 하는 경우도 있다. 다음 경우에 해당하는 예를 들어 보아라.

- Y가 0 혹은 1 의 값을 가지는 경우
- Y가 0이상의 양의 정수만을 갖는 경우

위 와 같은 경우 $EY|X$의 범위가 제한됨을 쉽게 알 수 있다. 하지만, $\mathrm{E}Y|X = X^T\boldsymbol{\beta}$ 의 방식으로 모형화를 하면 좌변과 우변의 범위가 다르게 나타나, 자연스러운 모형에서 벗어난다고 생각할 수 있다. 따라서 다음과 같이 조건부 평균의 변환된 값을 이용하여 모형화를 하기도 한다.

- $0 < E(Y|x) < 1$ for all x 이면,

  $$g(E(Y|x)) = x^\top\boldsymbol{\beta}$$

  where $g(z) = \log(z/1-z)$ for $z \in (0,1)$,

- $E(Y|x) > 0$ for all x 이면,

  $$g(E(Y|x)) = x^\top\boldsymbol{\beta},$$

  where $g(z) = \log(z)$ for $z > 0$.

사실 위의 경우 조건부 평균과 설명변수에 의한 모형의 범위를 맞춰주는 것 이상의 가정이 모형에 반영된다. 첫 번째의 경우 조건부 평균의 모형이 다음과 같이 쓰여지게 된다.

$$E(Y|x) = \frac{e^{x^\top\beta}}{1+e^{x^\top\beta}}$$

두 번째 모형의 경우에는 다음과 같이 쓰여질 것이다.

$$E(Y|x) = e^{x^\top\beta}$$

조건부 추정의 문제는 앞 서 살펴본 위험함수를 사용할 수 있다. (어떻게 하면 될까?)

- 위험함수 중에서는 제곱으로 만들어진 함수 외에도 우도함수가 개발되어 있다. (수리통계학에서 학습)

8.3 자료의 분위수 찾기

일차원 자료 X의 중심이 아니라 상위 25%, 혹은 하위 10%등 자료의 분위수에 관심이 있는 경우가 있다. 이 때 사용할 수 있는 모형에 대해서 알아보자. 여기서는 상위 25%에 대한 추정을 알아보도록 하자.

- $\mu^* \in \mathbb{R}$: 예상하는 τ의 위치 ($\Pr(X \leq \mu^*) = \tau$, $\tau \in (0,1)$) 라고 하자. 아래는 어떤 τ가 주어졌을때, μ^*를 찾는 방법에 대한 설명이다.

- $\tau(X-\mu)_+ + (1-\tau)(\mu - X)_+$: 데이터와 μ가 떨어진 크기를 제곱합수가 아닌 다음 함수로서 표현하자. ($z_+ = \max(z, 0)$)

- $\mathrm{E}(\tau(X-\mu)_+ + (1-\tau)(\mu - X))$를 최소화 하는 μ값은

  $$\frac{\partial}{\partial \mu}(\mathrm{E}(\tau(X-\mu)_+ + (1-\tau)(\mu - X)_+) = 0$$

  를 만족하는 μ를 찾으면 된다.

  $$\frac{\partial}{\partial \mu}(\mathrm{E}(\tau(X-\mu)_+ + (1-\tau)(\mu - X)_+) = -\tau \Pr(X-\mu > 0) + (1-\tau)\Pr(X - \mu \leq 0)$$

  이므로 정리하면

  $$\tau = \Pr(x \leq \mu)$$

즉,

$$\min_\mu (\mathrm{E}(\tau(X-\mu)_+ + (1-\tau)(\mu - X)_+)$$

를 이용하면, μ^*를 찾을 수 있다. (조건부 평균과 조건부 분위수를 찾는 방법을 비교해보자.)