

이 장에서는 데이터 분석을 시작하는 데 꼭 필요한 기초 선형대수를 공부한다. ~~선형대수~~는 데이터를 다루는 방법을 표시한 기호이자 언어다. ~~선형대수~~는 데이터를 다루는 법은 물론이고 연립방정식을 사용하여 미지수의 값을 구하는 법 등에 사용되는

② 데이터 분석의 기본적인 도구다. 따라서 선형대수에서 나오는 용어의 정의와 사용법에 익숙해져야 한다.

2.1 데이터와 행렬

선형대수(linear algebra)는 데이터 분석에 필요한 각종 계산을 돕는 학문이다. 데이터 분석을 하려면 수많은 숫자로 이루어진 데이터를 다루어야 한다. 하나의 데이터가 수십 개에서 수만 개의 숫자로 이루어져 있을 수도 있고, 또 이러한 데이터 수만 개가 하나의 집합을 이루고 있을 수도 있다.

선형대수를 사용하면 대량의 데이터를 포함하는 복잡한 계산 과정을 몇 글자 되지 않는 간단한 수식으로 서술할 수 있다. 따라서 데이터를 다루는 과정을 정확하고 간단하게 서술할 수 있다. 이를 위해 우선 선형대수에서 사용되는 여러 기호와 개념을 익혀보자.

각종 데이터를 선형대수의 기호로 나타낼 수 있다.

선형대수의 각종 연산 특히 행렬의 곱셈을 이용하여 데이터를 조작할 수 있다.

역행렬과 연립방정식의 해를 구하고 이를 회귀분석과 예측 문제에 활용할 수 있다.