

데이터 분석을 위한 선형대수학

2021년 8월 8일 일요일 오후 5:40



- 데이터들 간 유사성 판단
 - 행을 점으로 바꾼다 (vector)
 - 테이블 상 행끼리 비교한다.
 - 행끼리 거리를 측정
- 벡터
 - 여러 개의 숫자 모음
 - 벡터 공간의 원소
- Vector space = 집합
- Norm = 벡터의 길이
 - 원점으로부터 얼마나 떨어져 있나
- 벡터의 기본연산
 - 같은 차원의 벡터들 사이에서만 가능
- 벡터 간 거리
 - 데이터 사이의 유사성 측정 -> 거리함수
 - Manhattan Distance (L1)
 - basis의 차이/특성을 살리고 싶을때
 - Euclidean Distance (L2)
 - Cosine similarity
 - 데이터 간 구성차이 비교
- 장점, 단점? 어떤 상황에 무엇이 쓰이나?
- **Feature Space**
 - 주어진 데이터의 특징을 정의한 벡터공간
- **Feature vector**
 - Input vector중 필요 특징만 추출 한 것
- **Feature Engineering**
 - Input vector -> feature vector