😀

1. 벡터 공간과 부분 공간 (2)

PCA란?
👓 부분 공간
- 🖥️ 활용

👓 일차 결합
- 🖥️ 활용

👓 일차 독립 / 일차 종속
- 🖥️ 활용

👓 기저
- 🖥️ 활용

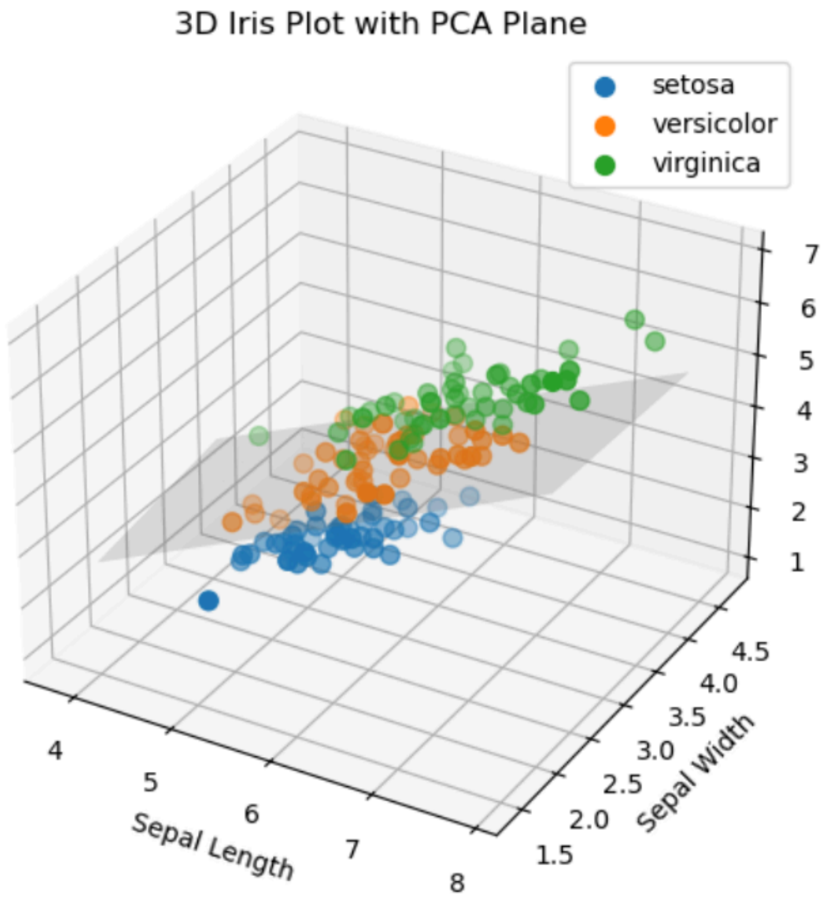
👓 직합
- 🖥️ 활용

👓 그람 슈미트 직교화
- 🖥️ 활용

👓 직교여공간
- 🖥️ 활용

PCA란?

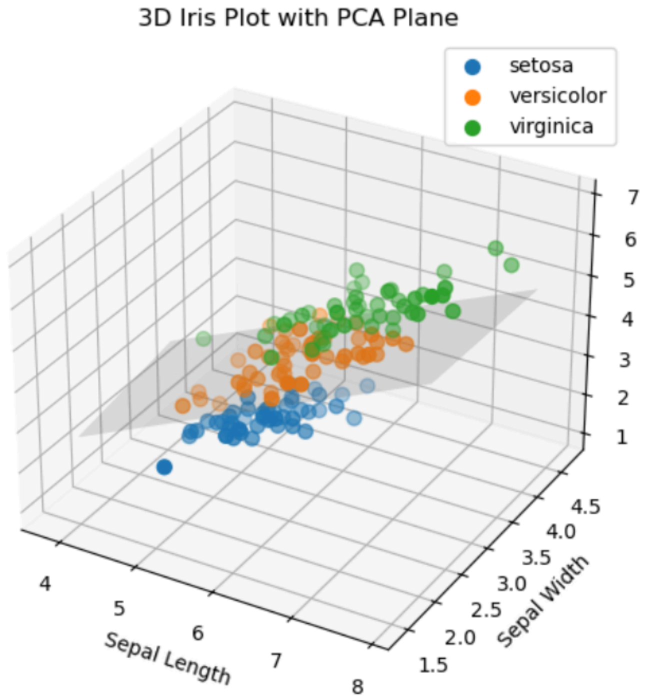
> 💡 전체 데이터 공간을 **정보가 많은 부분공간(주성분)**과 정보가 **적은 부분공간(여공간)**으로 직합 분해 하여 분리하는 것!



PCA를 통해 3차원 공간의 부분 공간인 2차원 공간으로 매핑되는 그림

🕶️ 부분 공간

> 💡 벡터 공간의 일부로, 덧셈, 스칼라곱 연산을 해도 닫혀 있는 공간.



PCA를 통해 3차원 공간의 부분 공간인 2차원 공간으로 매핑되는 그림

🖥️ 활용

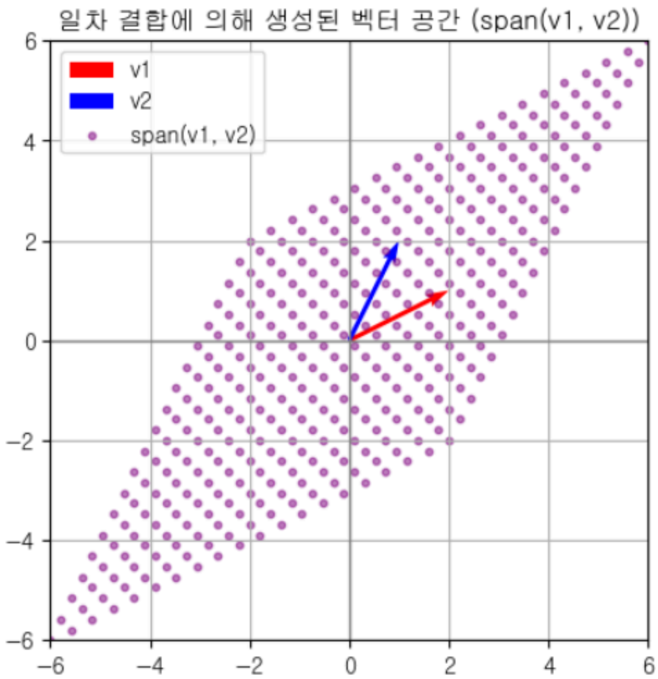
- PCA에서 선택되는 주성분이 span하는 공간이 부분 공간.
- 차원 축소, 특성 선택 시 의미 있는 정보만 남기는 과정 = 부분 공간에 투영하는 것.

👓 일차 결합

💡 벡터들에 실수의 곱을 한 뒤 모두 더하는 연산.

$\lambda_1 x_1 + \lambda_2 x_2 + \ldots$

공간의 확장과도 같은 의미!



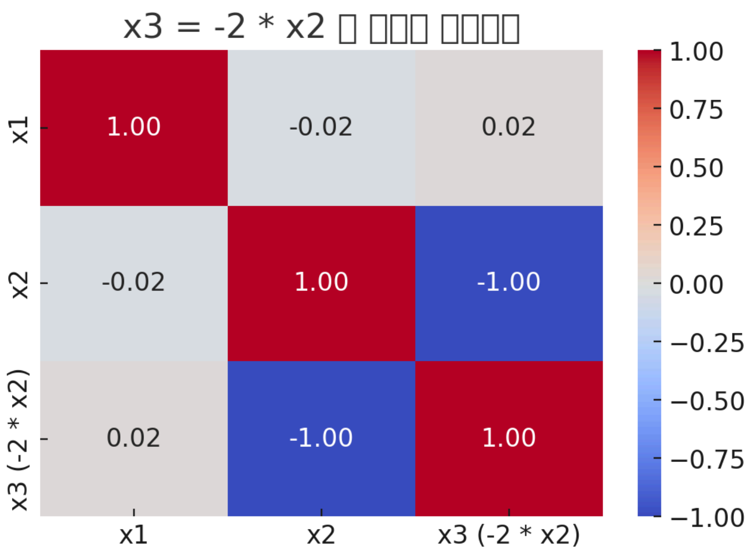
💻 활용

- 모든 데이터는 수많은 특성들의 가중합이 된다.
- 영화 추천 시스템 : 유저 취향과 영화 item 벡터의 선형 결합으로 모델링 가능.
  - 유저 취향 가중치 λ와 영화 벡터 x 사이의 선형결합 모델링.
  - $preference = \lambda_1 x_1 + \lambda_2 x_2 + ...$

👓 일차 독립 / 일차 종속

> 💡 $\lambda_1 x_1 + \lambda_2 x_2 + ... = 0$

> 💡 벡터들이 서로 중복되지 않는다 (방향이 서로 다르다) = **일차 독립**
>
> 중복되는 벡터가 존재한다 (방향이 서로 같다) = **일차 종속**

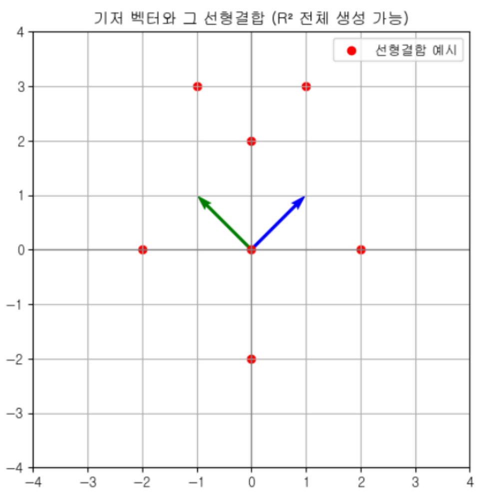


🖥️ 활용

- PCA 차원 축소
- 중복된 feature는 제거해야 한다. → 일차 종속인 벡터들을 찾아 제거한다.

👓 기저

> 💡 벡터 공간을 표현할 수 있는 최소한의 독립적인 벡터 집합. (4차원의 기저 벡터는 4개)
>
> **정규직교기저** : 모든 벡터가 서로 직교하고 크기가 1인 정규벡터.

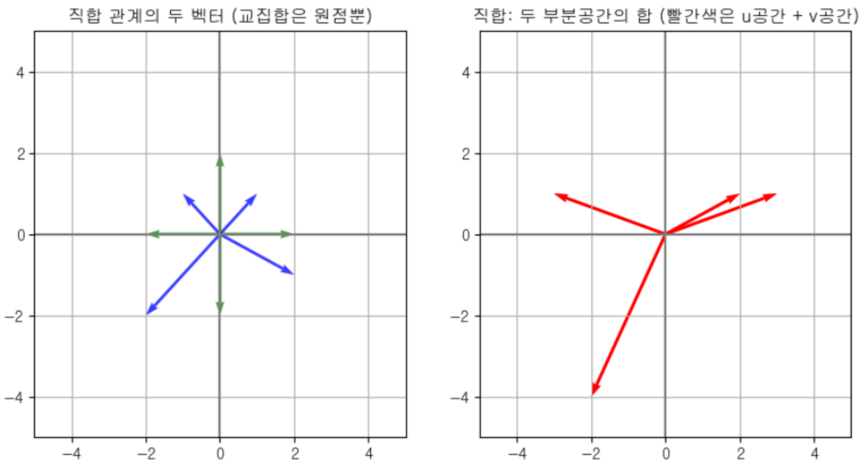


🖥️ 활용

- 데이터 압축 : 데이터의 차원을 축소하면 기저의 수도 줄어든다.
- PCA - 고차원 데이터에서 새로운 기저 벡터로 좌표 재구성.

👓 직합

> 💡 원점을 제외하고 서로 겹치지 않는 두 부분공간(W_1, W_2)을 합쳐 하나의 공간이 되는 경우.

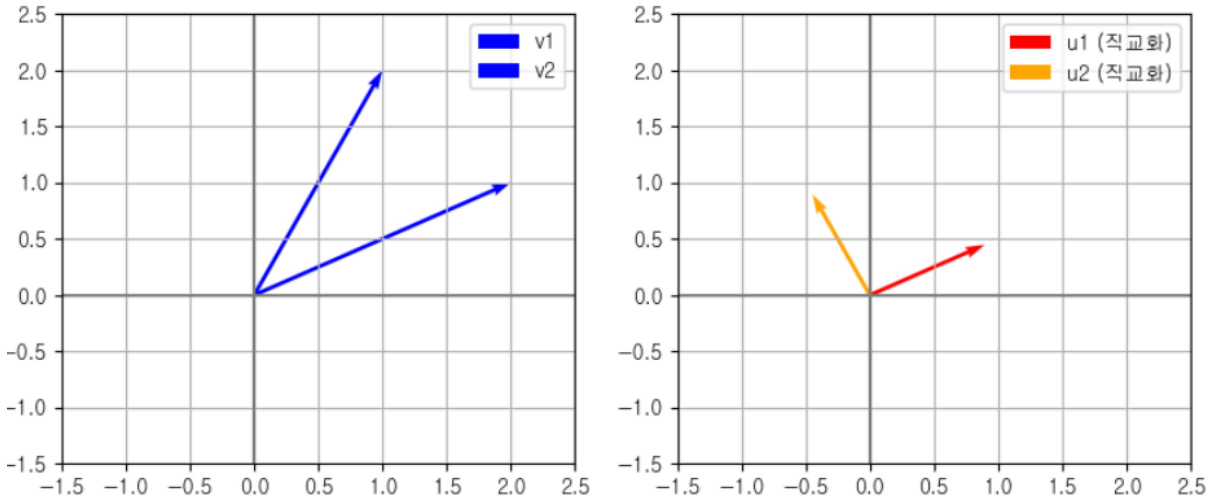


🖥️ 활용

- PCA : 원본 데이터 = 주성분 공간 ⊕ 여공간
  - 주성분 공간 : 정보가 많은 축의 데이터 공간 (남겨야 하는 대상)
  - 여공간 : 정보가 적고 덜 중요한 축의 데이터 공간 (제거 대상)

👓 그람 슈미트 직교화

> 💡 서로 직교하는 벡터로 벡터 집합을 변환 시키는 과정

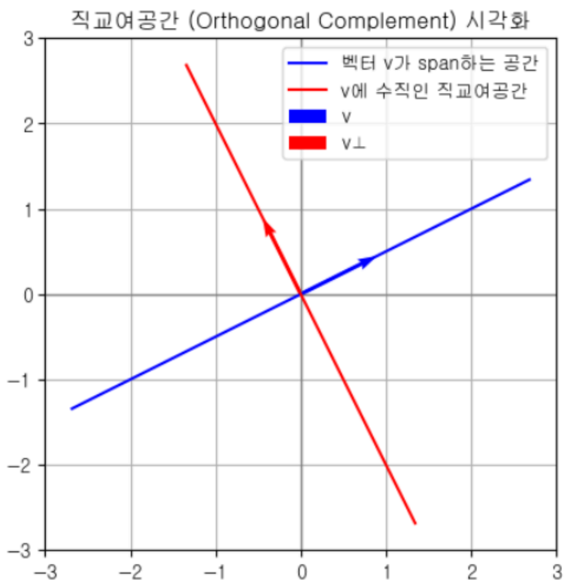


💻 활용

- 데이터에서 중복/상관관계를 제거하기 위한 전처리로 사용 가능. (계산이 단순해지고 해석이 쉬워짐)

👓 직교여공간

> 💡 어떤 부분공간과 수직인 벡터들로 이뤄진 공간



💻 활용

- PCA의 주성분공간과 여공간은 서로 직교여공간!