

## ✧ PCA의 특징

- 새로운 축은 독립이며 직각임
- 원본 데이터의 차원이  $p$ 라면 새롭게 만들어진 데이터의 차원은  $k(k < p)$
- 원본 데이터  $X$ 는  $U$ 와  $V$ 의 곱으로 분리됨
- $U$ 는 데이터를 새로운 차원  $k$ 로 설명함
- $V$ 는 원본 차원과 축소 차원의 관계를 설명함
- 정보를 많이 잃어버리지 않고 차원을 축소시킴

## ✧ DBSCAN 용어

- 이웃 벡터 : 반경  $\epsilon$  안에 포함된 데이터들
- 핵심 벡터 :  $n$ 개 이상의 이웃 벡터를 가짐
- 직접 접근 가능 : 핵심 벡터와 이웃 벡터와의 관계( $p \rightarrow q$ )
- 접근 가능 : 연속적으로 이루어짐
  - 핵심 벡터  $\rightarrow$  이웃 벡터  $\rightarrow$  이웃 벡터  $\rightarrow$  이웃 벡터 ... 일 때에 접근 가능으로 표현( $p \Rightarrow q$ )
  - 연결된  $p$ 와  $q$ 사이에 접근 가능한 벡터가 있다면  $p \Leftrightarrow q$ 로 표현

## ✧ DBSCAN vs k-means

- 밀도기반 vs 거리기반
- 어떤 형태의 군집도 잘 잡는 편 vs 원이나 구 모양에 최적화되어 있음
- 노이즈가 정의됨 vs 노이즈가 정의가 안됨
- 직관적 vs 수학적
- 프로그래밍으로 구현 vs 컴퓨터가 없어도 계산으로 풀 수 있음

## ✧ PCA에서의 새로운 차원 $k$

- 고윳값 분해를 하여 가장 작으면서 분산을 많이 설명할 수 있는 상위  $k$ 개의 추상적인 축을 선택