

Chapter08

주성분과 요인 분석

I 주성분 활용 (PCA)

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

- 차원 축소.
- 고차원 클러스터의 시각화.

I 주성분 분석: 차원 축소

- 주성분을 새로운 좌표축으로 사용할 수 있다.
- 주성분의 개수는 원래 차원의 개수와 같다.
- 그런데 주성분은 분산의 크기 순서대로 정렬되어 있다.

$$\rightarrow \sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \dots$$

- 그러므로, 분산이 작은 순서대로 차원을 축소해 나갈 수 있다.

I 주성분 분석: 차원 축소의 장단점

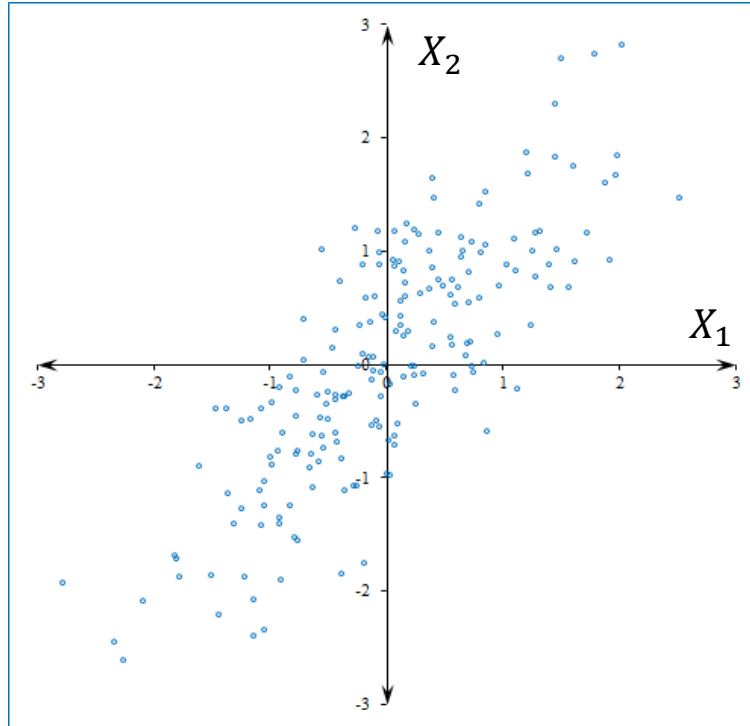
장점	단점
<div><div>✓ 가장 뚜렷한 특징만 뽑아냄.</div><div>✓ 데이터 표현의 간소화.</div><div>✓ 연산, 메모리 부하 감소.</div></div>	<div><div>✓ 해석이 어려움.</div><div>✓ 작은 디테일 손상.</div></div>

I 주성분 분석: 차원 축소 결과

- Reduced dimension input: 차원축소 후의 데이터 좌표.
- Loading: 정규화된 주성분.
- Communality: 차원축소 후 원래의 변수가 어느정도 복원 되었는가.

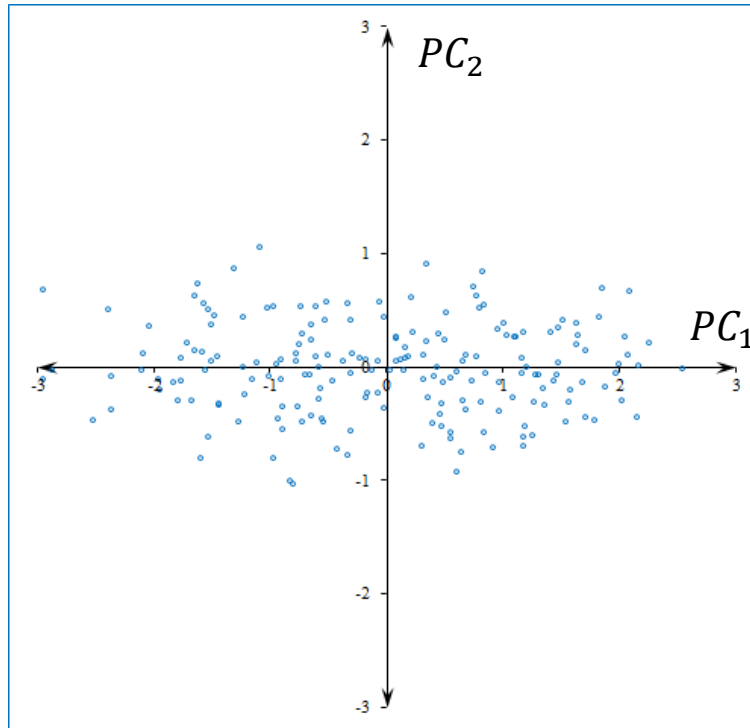
I 주성분 분석: 차원 축소의 원리

- 다음과 같이 데이터 좌표가 분포되어 있다고 가정한다.



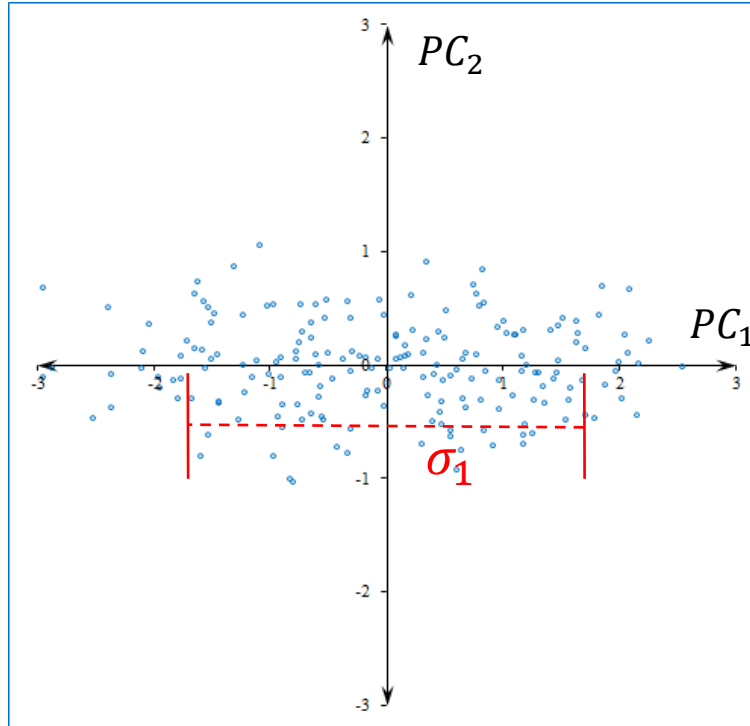
I 주성분 분석: 차원 축소의 원리

- 주성분 PC_1 과 PC_2 를 새로운 좌표축으로 사용하였을 때.



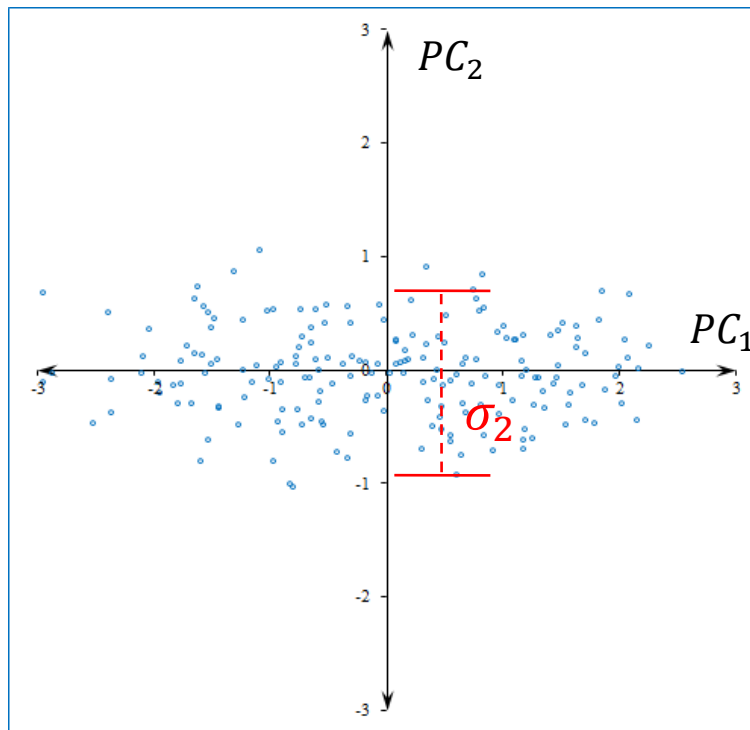
I 주성분 분석: 차원 축소의 원리

- PC_1 는 변동이 가장 큰 방향.



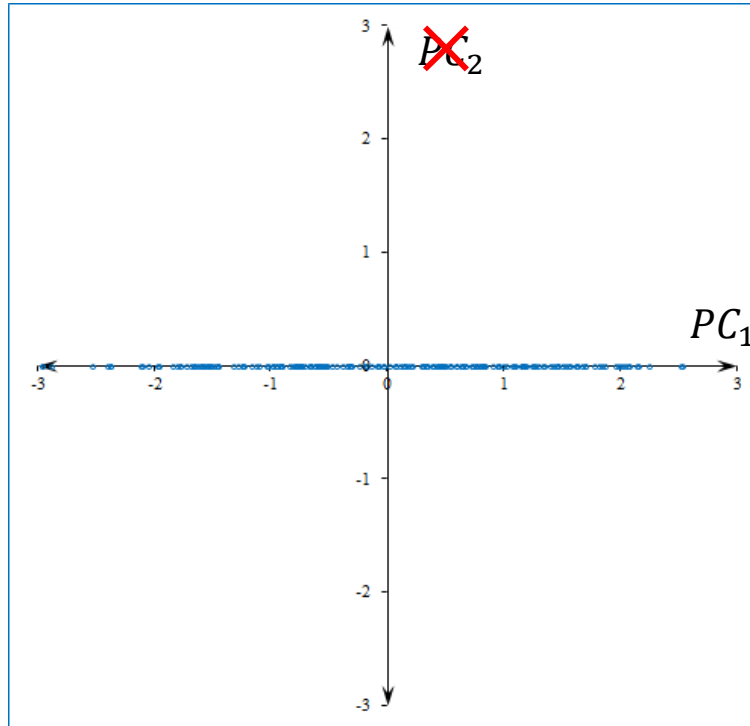
I 주성분 분석: 차원 축소의 원리

- PC_2 는 변동이 상대적으로 작은 방향.



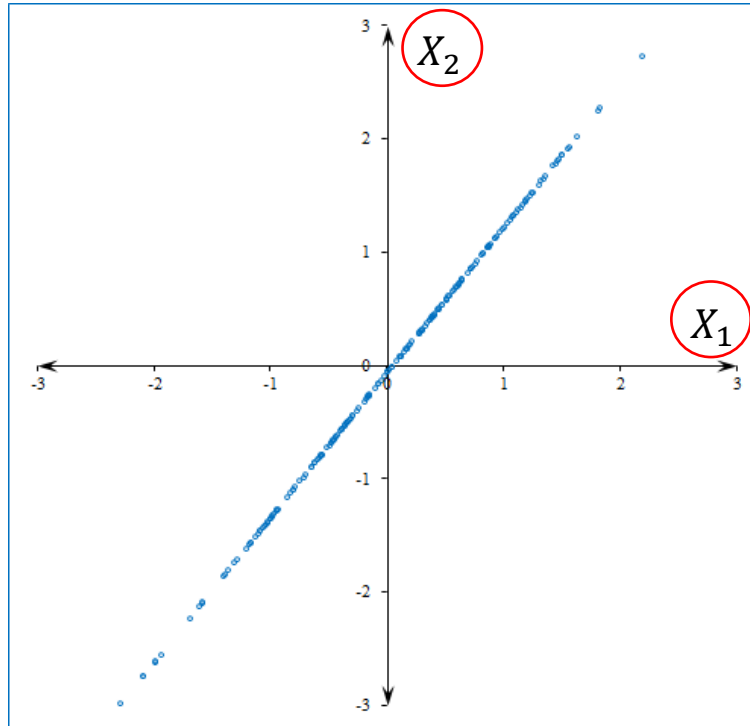
I 주성분 분석: 차원 축소의 원리

- PC_2 방향으로 차원 축소.



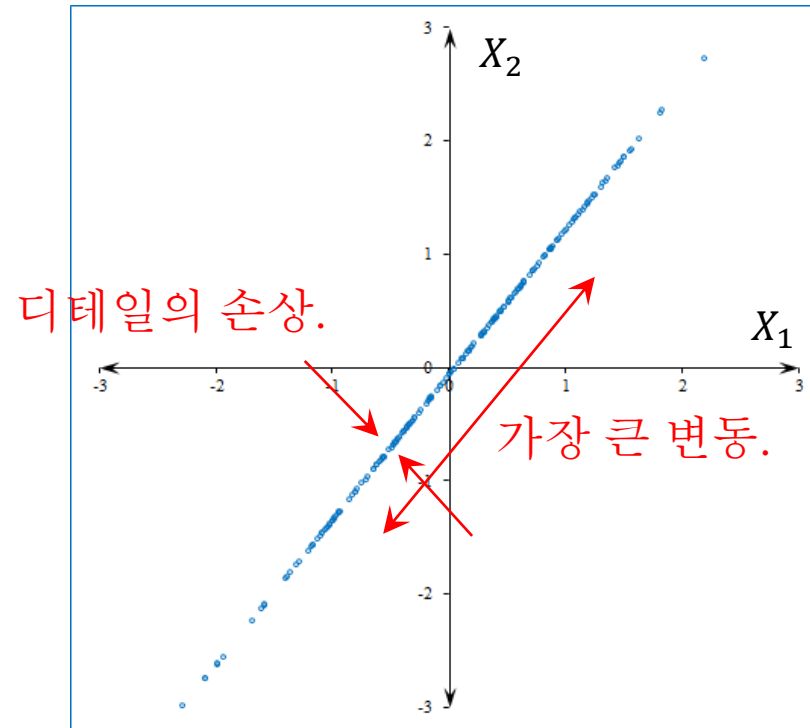
I 주성분 분석: 차원 축소의 원리

- 원 좌표축으로 돌아와 본다 (reduced dimension input).



I 주성분 분석: 차원 축소의 원리

- 원 좌표축으로 돌아와 본다 (reduced dimension input).

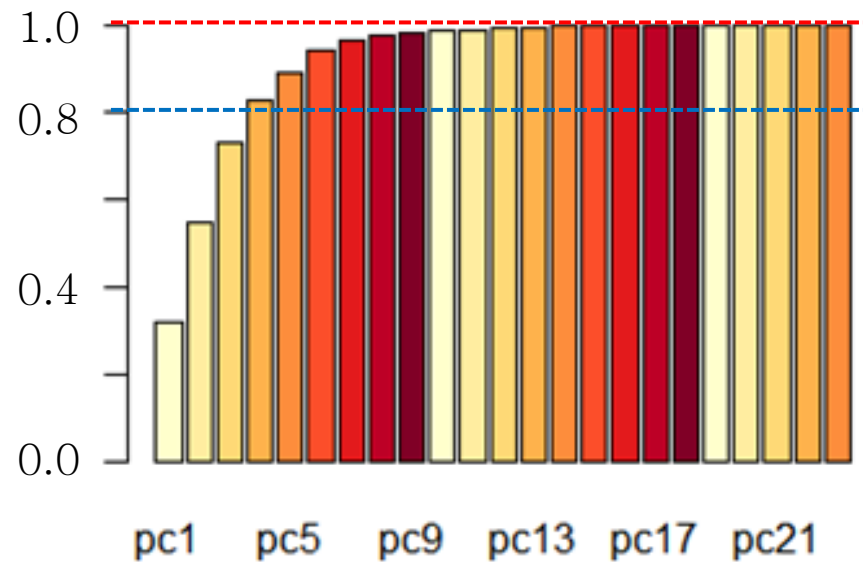


I 주성분 분석: 차원 축소 기준

- 전체 분산은 개개 주성분 방향의 분산의 합으로 구할 수 있다.

$$\sigma_{total}^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots$$

- 분산에 목표치를 두고 첫 몇개의 주성분만 남겨둔다.



I 주성분 분석: 고차원 클러스터의 시각화 응용

- 고차원 클러스터를 평면 (2D)에 투영하여 시각화하려고 한다.
- 분산이 큰 순서대로 두개의 주성분 (PC_1 과 PC_2)을 사용한다.
- 이 두 주성분은 새로운 평면을 정의한다.
- 고차원 데이터 좌표를 PC_1 과 PC_2 평면상의 좌표로 투영한다.

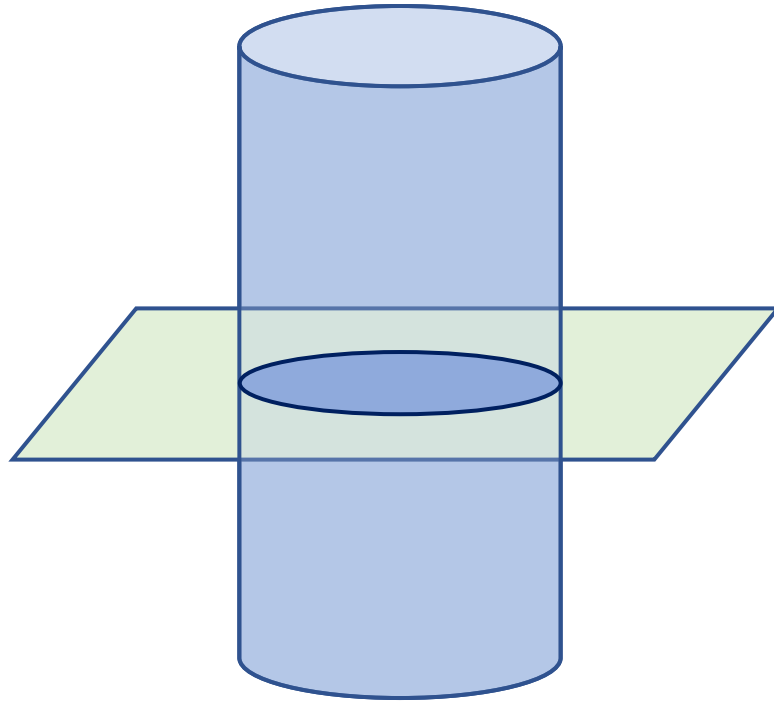
I 주성분 분석: 고차원 클러스터의 시각화 응용

- 데이터 좌표가 실린더 형상으로 분포해 있다고 가정해 본다.



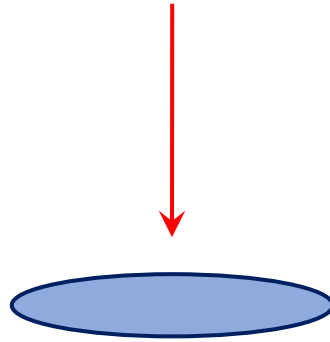
I 주성분 분석: 고차원 클러스터의 시각화 응용

- 2D 단면을 다음과 같이 자른다.



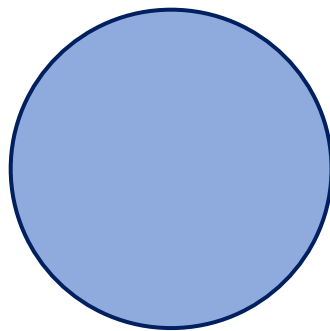
I 주성분 분석: 고차원 클러스터의 시각화 응용

- 단면을 정면에서 바라 본다.



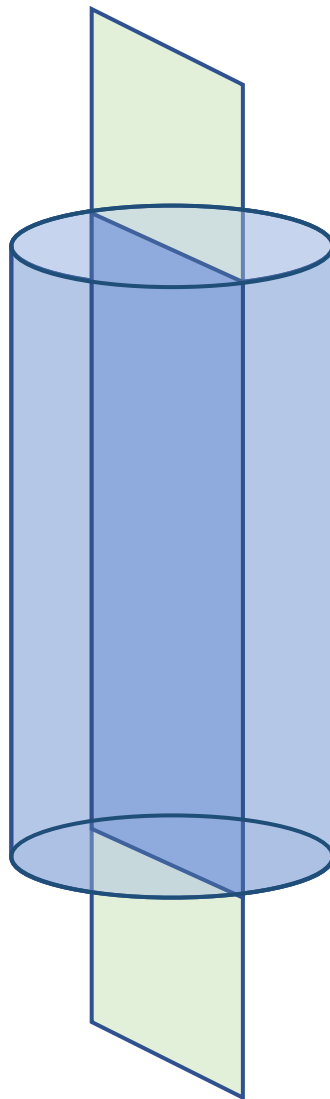
I 주성분 분석: 고차원 클러스터의 시각화 응용

- 그러면 이렇게 보이겠죠.



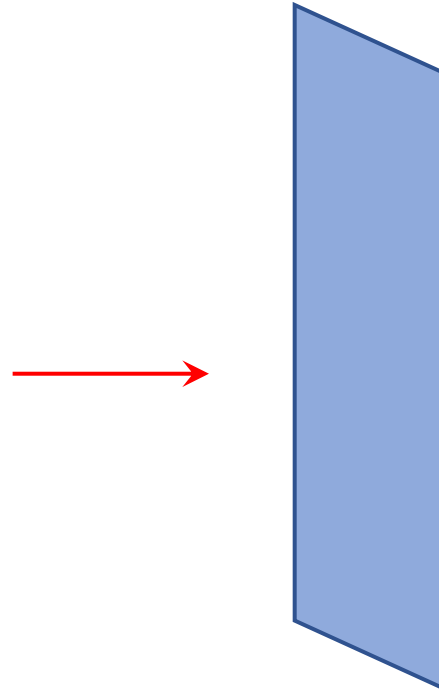
I 주성분 분석: 고차원 클러스터의 시각화 응용

- 또다른 2D 단면을 자른다.



I 주성분 분석: 고차원 클러스터의 시각화 응용

- 단면을 정면에서 바라 본다.

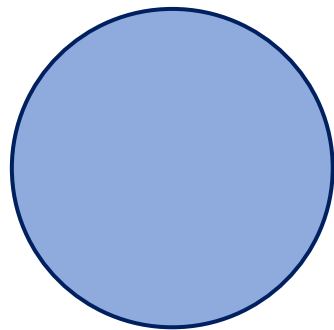


I 주성분 분석: 고차원 클러스터의 시각화 응용

- 그러면 이렇게 보이겠죠.



I 주성분 분석: 고차원 클러스터의 시각화 응용



대



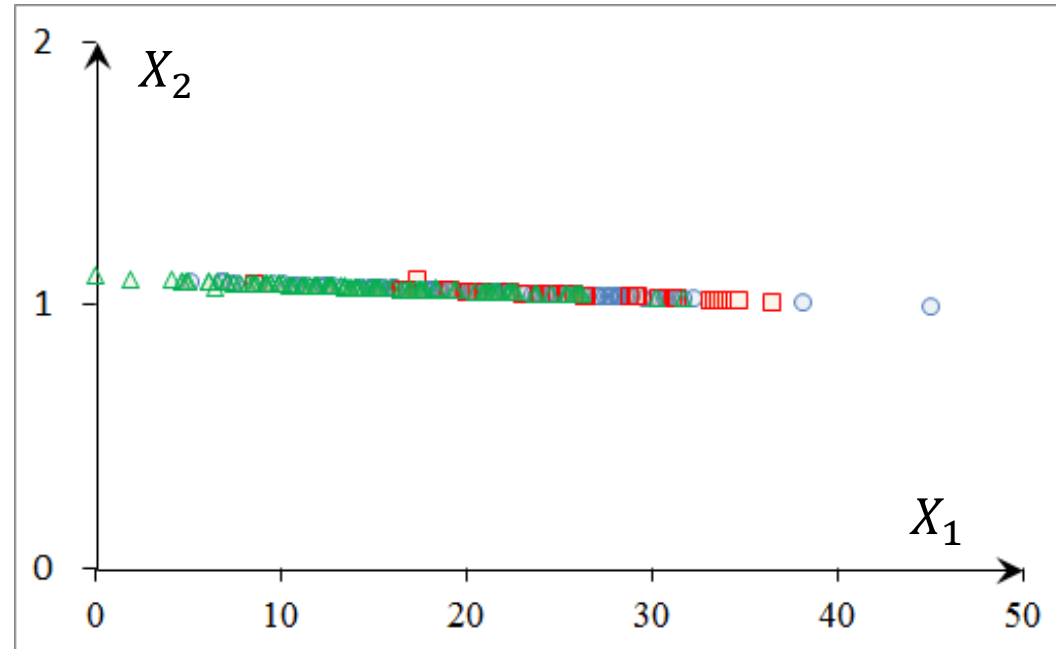
- ✓ 더 넓게 퍼져있는 단면.
- ✓ 투영하기에 좋음.

I 주성분 분석: 고차원 클러스터의 시각화 응용

- 단면을 잘 선택하면 시각화에 유리하다.
- PC_1 과 PC_2 는 가장 넓게 퍼져있는 단면을 정의한다.

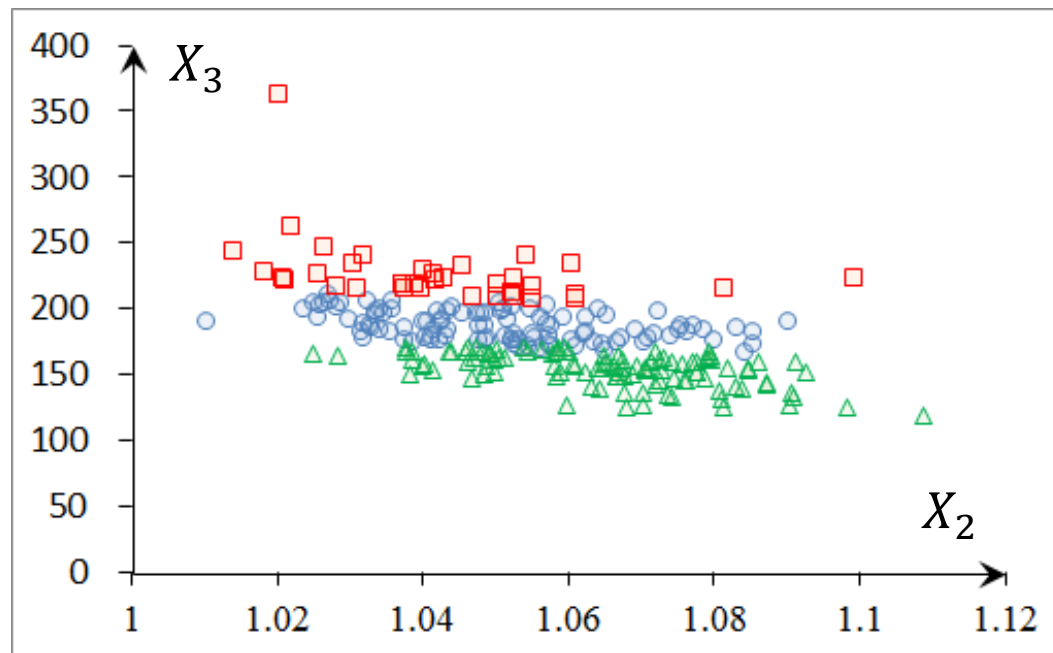
I 주성분 분석: 고차원 클러스터의 시각화 예

- 원 좌표 X_1 과 X_2 사용.



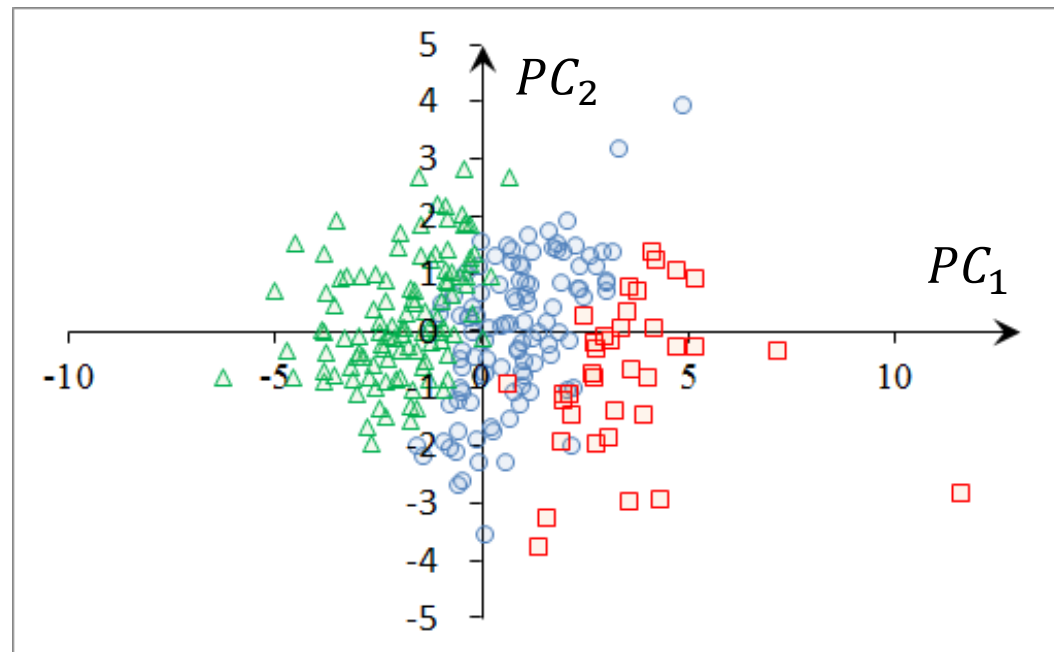
I 주성분 분석: 고차원 클러스터의 시각화 예

- 원 좌표 X_2 과 X_3 사용.



I 주성분 분석: 고차원 클러스터의 시각화 예

- PC_1 과 PC_2 로 정의되는 단면에 투영.



I 끝.

감사합니다.

