Chapter. 02

선형대수학 응용

주성분 분석의 해석

FASTCAMPUS
ONLINE
금융공학/퀀트

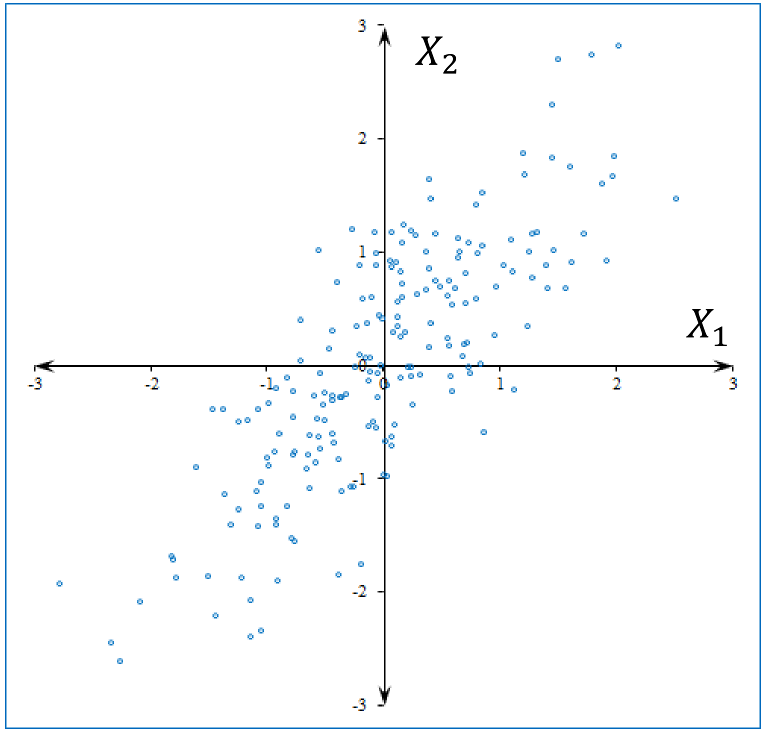
강사. 장순용

I 키포인트

- 고유값 분해와 주성분.
- 특이값 분해와 주성분.
- 고유값 분해와 특이값 분해의 관계.

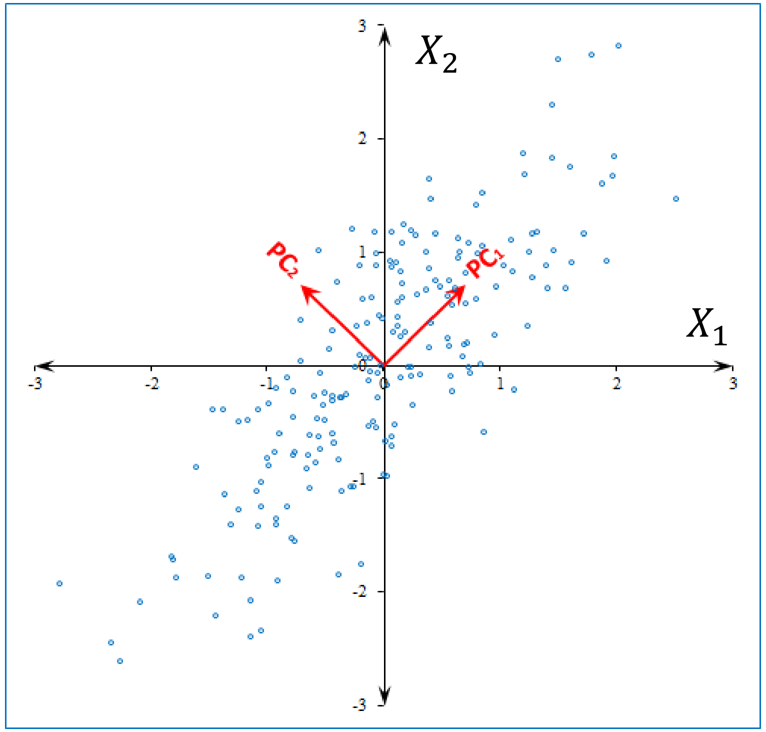
I 주성분: 복습

- 다음과 같이 분포된 2차원 데이터로 주성분에 대해서 알아본다.



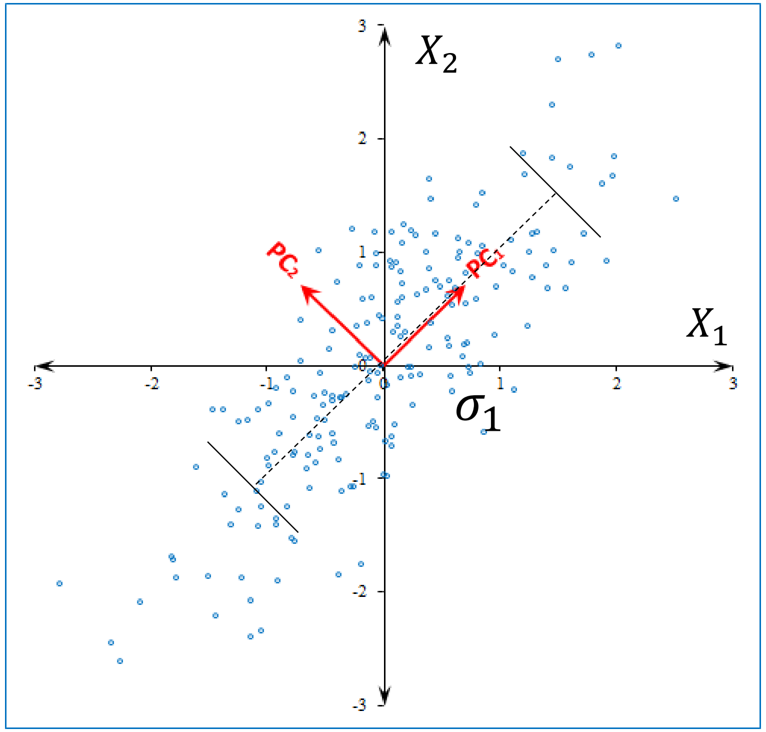
I 주성분: 복습

- PC_1과 PC_2는 서로 직교함.



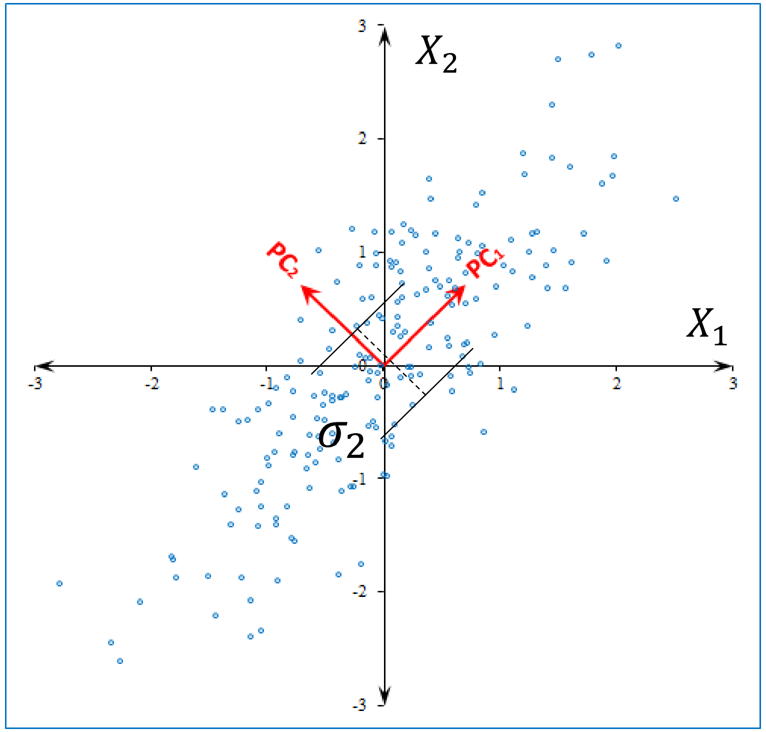
l 주성분: 복습

- PC_1에 해당하는 변동은 σ_1으로 나타낼 수 있다. 가장 큰 변동에 해당한다.



I 주성분: 복습

- PC_2에 해당하는 변동은 σ_2이며 $\sigma_2 < \sigma_1$이다.



I 고유값 분해와 주성분

- $\boldsymbol{X}$가 centering된 관측값으로 이루어진 행렬이라면, 분산공분산행렬에 비례하는 행렬 $\boldsymbol{M}$은 다음과 같이 구할 수 있다.

$$\boldsymbol{M} = \boldsymbol{X}\,\boldsymbol{X}^{t}$$

- 이제는 행렬 $\boldsymbol{M}$에 고유값 분해를 적용하여 $\boldsymbol{M} = \boldsymbol{Q}\,\boldsymbol{\Lambda}\,\boldsymbol{Q}^{t}$과 같이 표현한다.

$\Rightarrow$ $\boldsymbol{Q}$의 개개 컬럼은 서로 직교하는 주성분 (PC)이 된다.

$\Rightarrow$ $\boldsymbol{\Lambda}$의 대각선 원소는 개개 주성분의 분산 σ_i^2에 해당된다.

I 특이값 분해와 주성분

- $\boldsymbol{X}$가 centering된 관측값으로 이루어진 행렬이라면 이것에 특이값 분해를 적용하여 $\boldsymbol{X} = \boldsymbol{U\Sigma V}^t$과 같이 표현한다.

$\Rightarrow$ $\boldsymbol{U}$의 개개 컬럼은 서로 직교하는 주성분 (PC)이 된다.

$\Rightarrow$ $\boldsymbol{\Sigma}$의 대각선 원소는 개개 주성분의 표준편차 σ_i에 해당된다.

l 고유값 분해와 특이값 분해의 관계

- $\boldsymbol{X}$가 centering된 관측값으로 이루어진 행렬이라면 이것에 특이값 분해를 적용하여 $\boldsymbol{X} = \boldsymbol{U\Sigma V^t}$과 같이 표현한다.

- 분해된 $\boldsymbol{X}$로 분산공분산행렬을 표현한다.

$$\boldsymbol{M} = \boldsymbol{X}\,\boldsymbol{X^t} = \boldsymbol{U\Sigma V^t}(\boldsymbol{U\Sigma V^t})^{\boldsymbol{t}}$$

$$= \boldsymbol{U\Sigma V^t}\,\boldsymbol{V\Sigma^t U^t}$$

$$= \boldsymbol{U\Sigma\Sigma^t U^t}$$

직교성 $\boldsymbol{V^t V} = \boldsymbol{I}$

$\Rightarrow$ $\boldsymbol{\Sigma\Sigma^t}$는 분산 σ_i^2을 원소로 갖는 대각행렬이다. 고유값 행렬 $\boldsymbol{\Lambda}$와 같다는 것이다.

$\Rightarrow$ 그리고 왼쪽 특이벡터 행렬 $\boldsymbol{U}$가 바로 고유벡터 행렬 $\boldsymbol{Q}$의 역할을 한다.

׀ 고유값 분해와 특이값 분해의 관계

- 그러면 다음과 같이 정리해 본다.

고유값 분해	특이값 분해
$X\,X^t = Q\,\Lambda\,Q^t$	$X = U\Sigma V^t$ $X\,X^t = U\Sigma\Sigma^t U^t$
$\Lambda = \Sigma\Sigma^t$	
$Q = U$	

I 끝.

Life Changing Education

감사합니다.

FAST CAMPUS
ONLINE
장순용 강사.