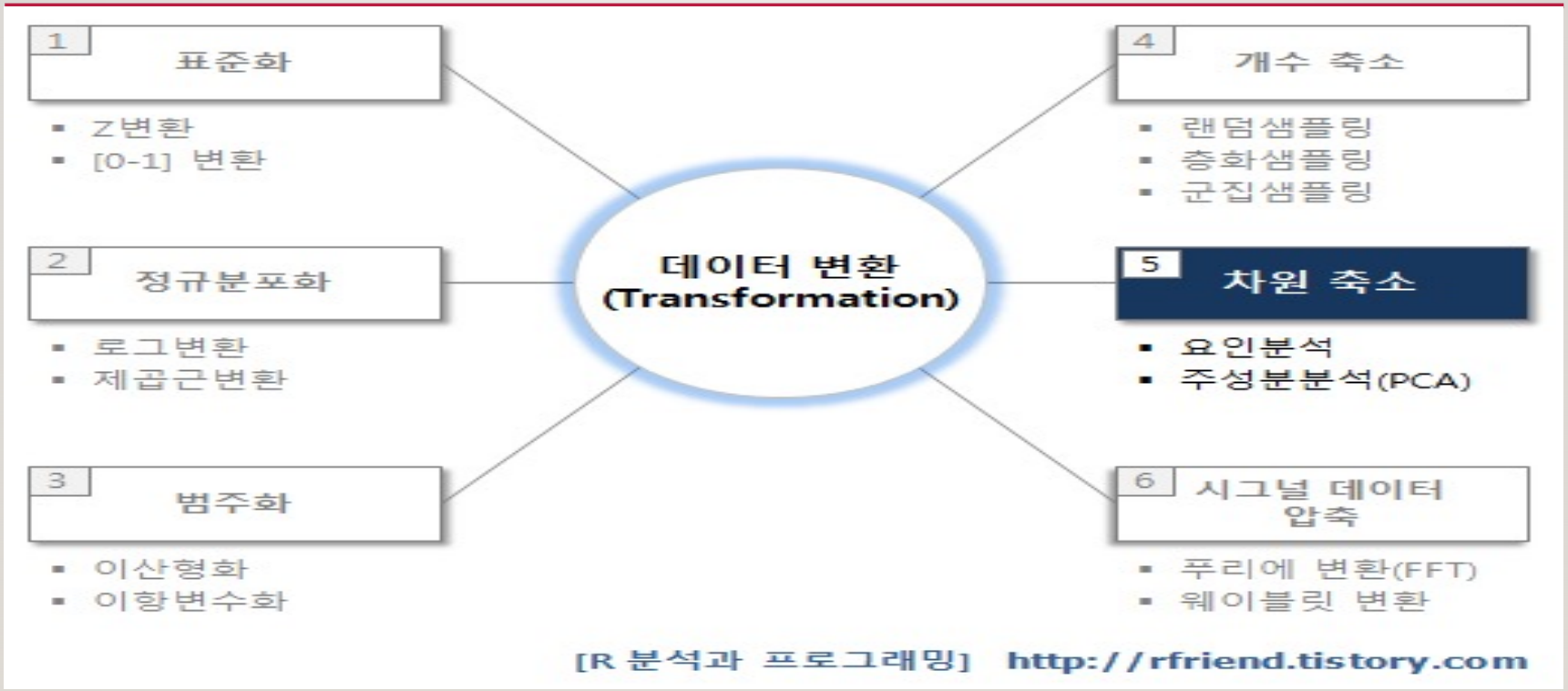
# PCA

21500468 이규석

# 목차

- 서론: 데이터 변환
- 본론: PCA란?
- 결론: 정리

# 서론: 데이터 변환


데이터 변환
(Transformation)
1 표준화
▪ Z변환
▪ [0-1] 변환
2 정규분포화
▪ 로그변환
▪ 제곱근변환
3 범주화
▪ 이산형화
▪ 이항변수화
4 개수 축소
▪ 랜덤샘플링
▪ 층화샘플링
▪ 군집샘플링
5 차원 축소
▪ 요인분석
▪ 주성분분석(PCA)
6 시그널 데이터 압축
▪ 푸리에 변환(FFT)
▪ 웨이블릿 변환
[R 분석과 프로그래밍] http://rfriend.tistory.com

# 본론: PCA란?

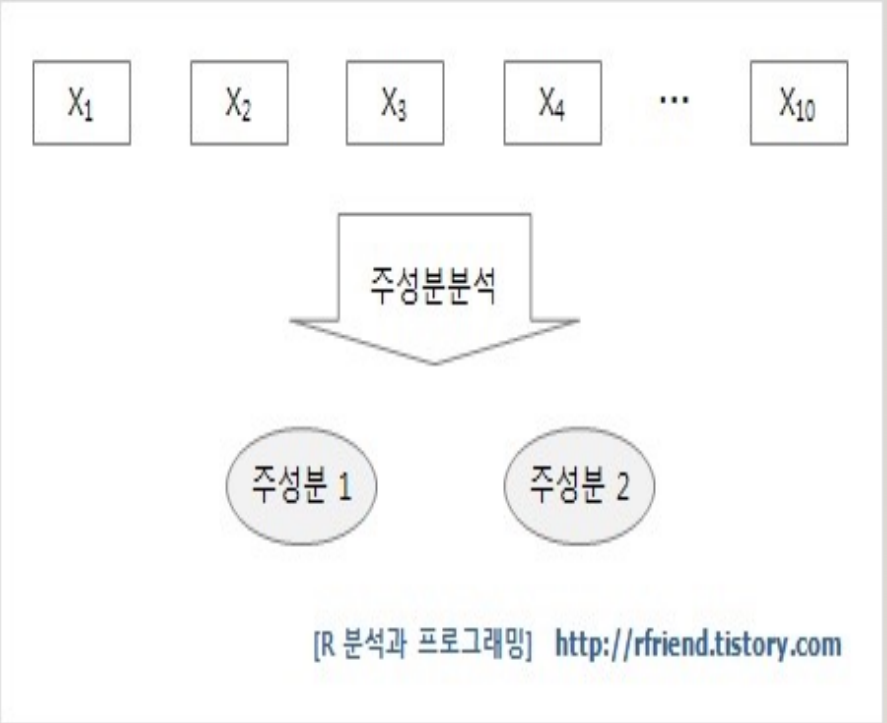
- Principal Component Analysis (주성분 분석)

- 상관성이 높은 여러 변수들의 선형 조합으로 새로운 변수를 생성 (첫번째 주성분)
- 두번째 주성분으로는, 첫번째 주성분과는 상관성이 가장 낮은 형태로 조합
- 첫번째 주성분으로 설명되지 못하는 나머지 변동을 가장 잘 설명

# 본론: PCA란?

- PCA 왜 사용할까?

1) 소수의 주성분만 사용가능
2) 연산 속도 개선



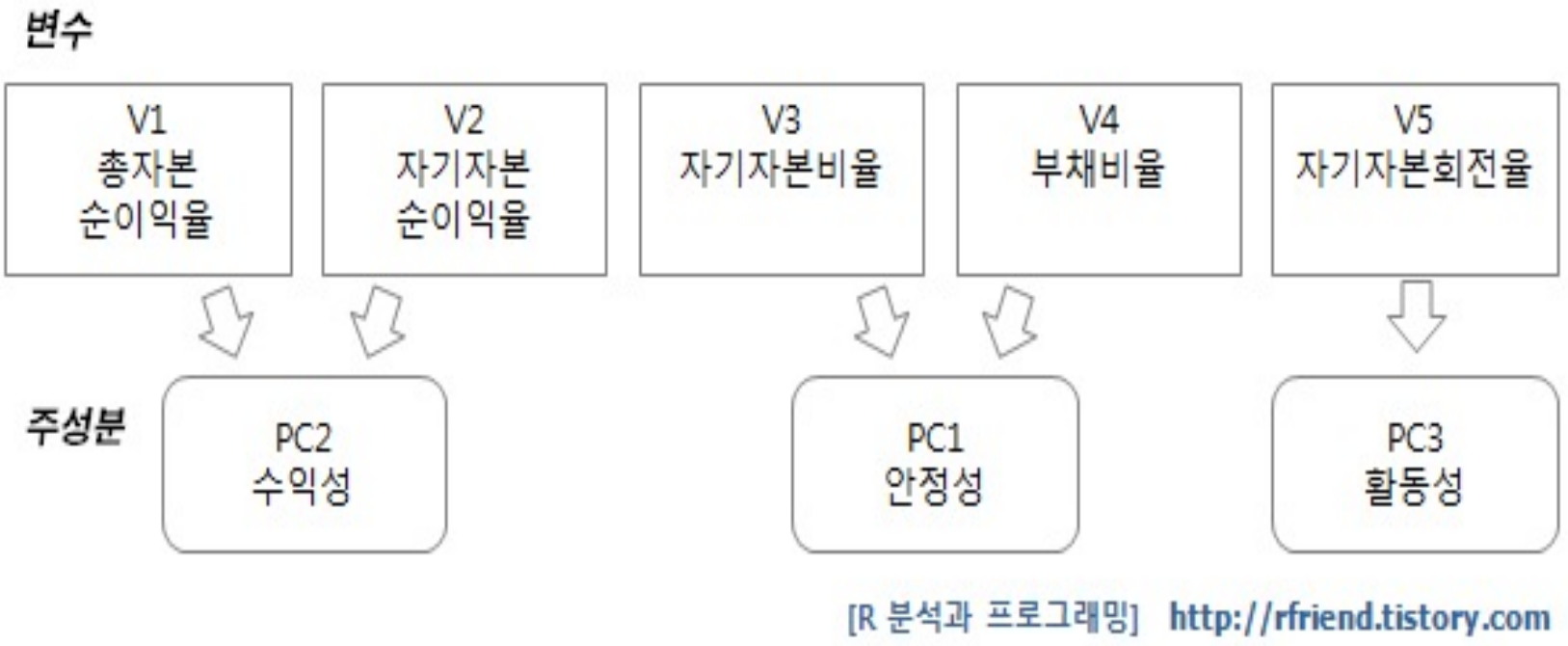
## 본론: PCA란?

# R 프로그래밍

데이터: 국내 증권회사의 주요 재무제표' (2007.3.31 기준)

# 본론: PCA란?

R 프로그래밍



# 본론: PCA란?

## R 프로그래밍

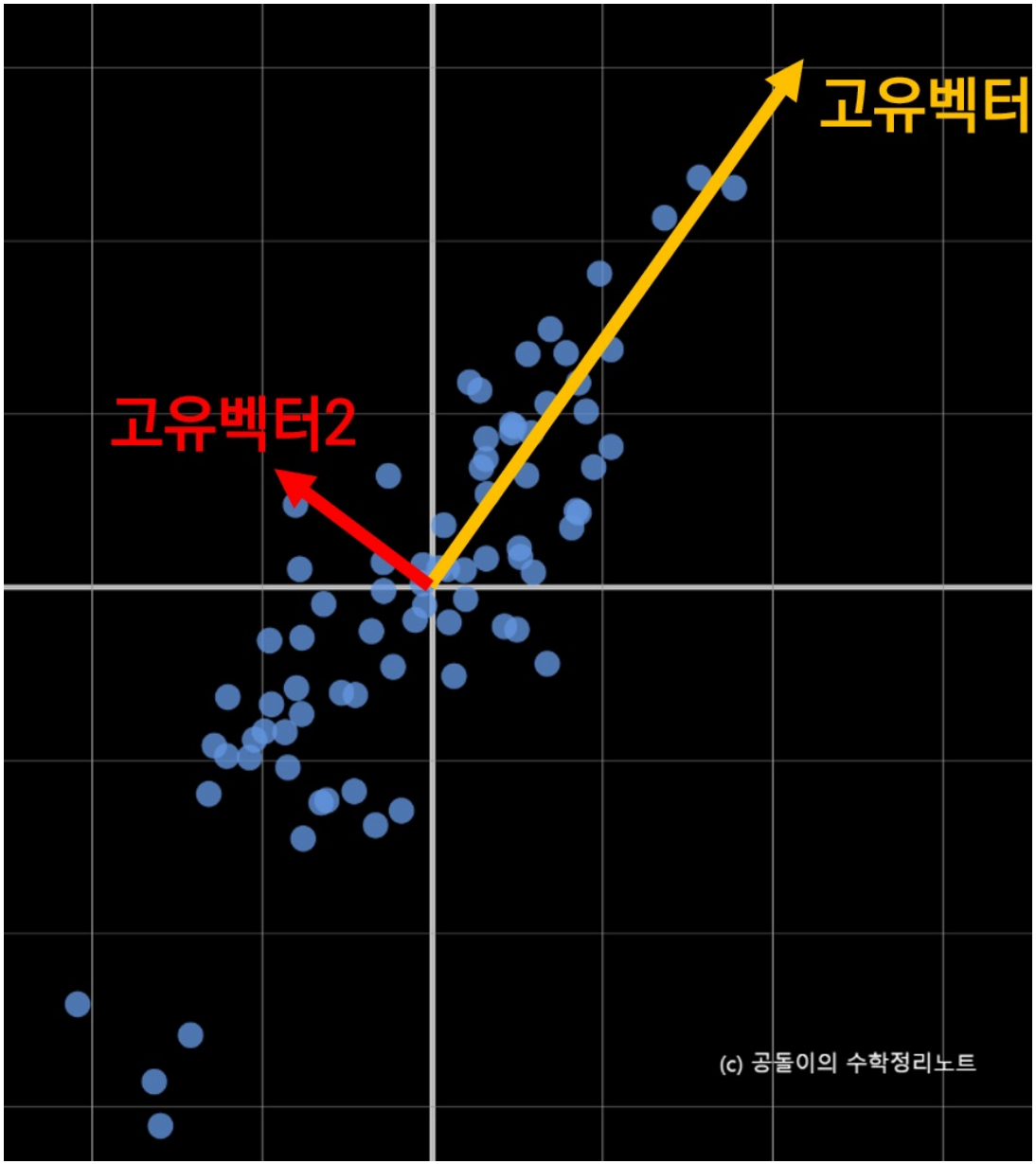
- **PC1** = 0.076*V1_s - 0.394*V2_s + **0.569*V3_s** + **0.559*V4_s2** - 0.447*V5_s )
- **PC2** = **-0.779*V1_s - 0.565*V2_s** - 0.162*V3_s - 0.196*V4_s2 - 0.086*V5_s )

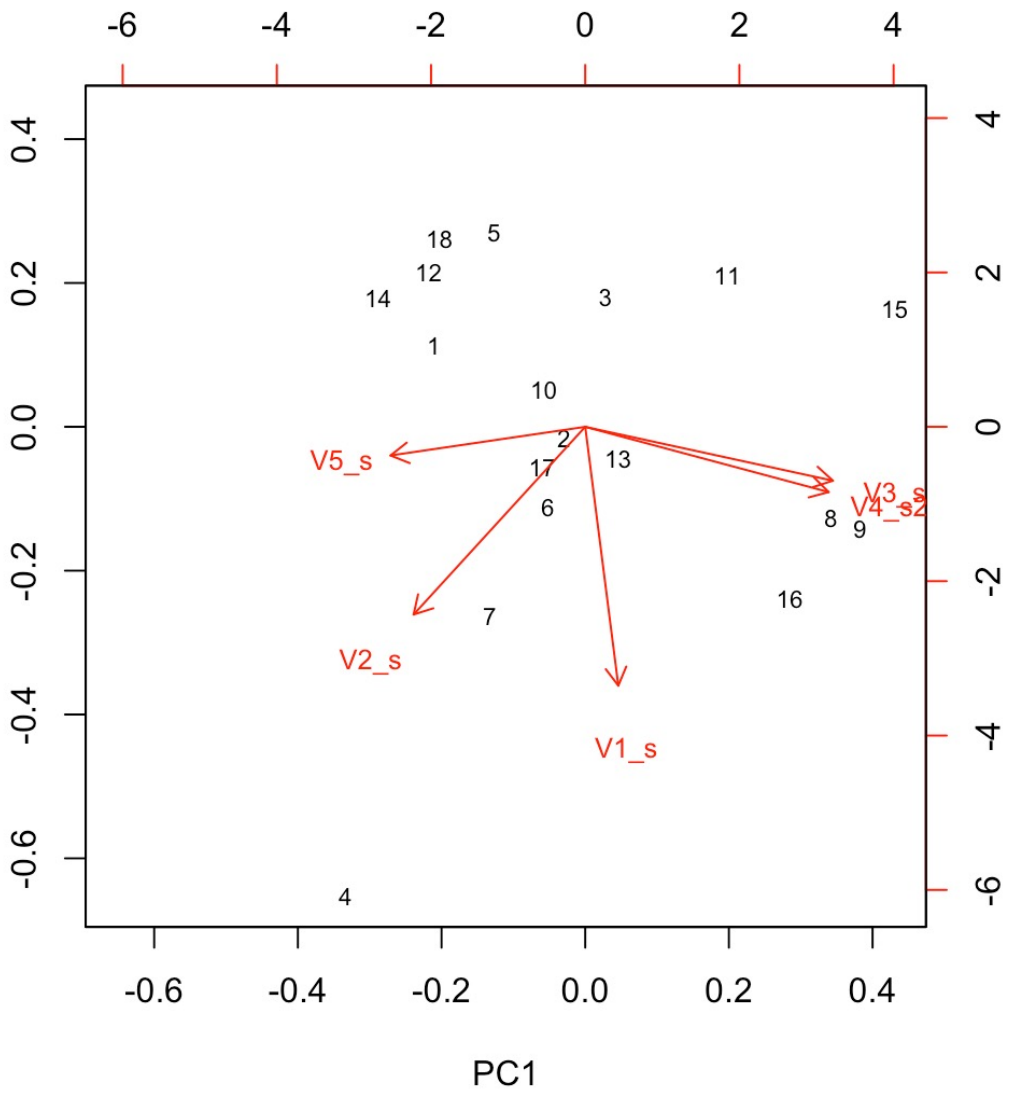
```
Standard deviations (1, .., p=5):
[1] 1.6617648 1.2671437 0.7419994 0.2531070 0.1351235

Rotation (n x k) = (5 x 5):
             PC1         PC2          PC3          PC4         PC5
V1_s   0.07608427 -0.77966993  0.0008915975 -0.140755404  0.60540325
V2_s  -0.39463007 -0.56541218 -0.2953216494  0.117644166 -0.65078503
V3_s   0.56970191 -0.16228156  0.2412221065 -0.637721889 -0.42921686
V4_s2  0.55982770 -0.19654293  0.2565972887  0.748094314 -0.14992183
V5_s  -0.44778451 -0.08636803  0.8881182665 -0.003668418 -0.05711464
```

# QUESTION: 몇개의 PCs를 사용할 것인가?

- There is no universal rules,  it uses rule of thumb.

1) 누적 기여율(설명된 분산의 누적 비율) 최소 (at least) 0.8이상
2) 평균 분산보다 큰 PC만 선별하기(표준화한 데이터에 대한 상관관계행렬을 사용할 경우, 고유값이 최소 1보다 큰 PC)
3) Screen Plot을 그려봤을 때, 꺾이는 부분(elbow)이 있다면 elbow 지점 앞의 PC 개수 선택.


고유벡터
고유벡터2
(c) 공돌이의 수학정리노트


-6
-4
-2
0
2
4
0.4
0.2
0.0
-0.2
-0.4
-0.6
18
5
12
11
14
3
15
1
10
2
V5_s
17
13
V3_s
6
V4_s
8
9
16
7
V2_s
V1_s
4
-0.6
-0.4
-0.2
0.0
0.2
0.4
PC1

# 결론: 정리

- PCA를 통해 주성분만으로, 데이터 분석이 가능하다.
- 이러한 Dimension Reduction 기법이 현재 딥러닝 연구 분야에서도 활발히 이뤄지고 있다.
- 선형대수학을 열심히 공부해야겠다.

# 감사합니다

참고자료: HTTPS://RFRIEND.TISTORY.COM/61

HTTPS://ANGELOYEO.GITHUB.IO/2019/07/27/PCA.HTML