

주성분분석

이재용, 임요한

서울대학교
통계학과

2017년 8월

노트. 다루는 내용 I

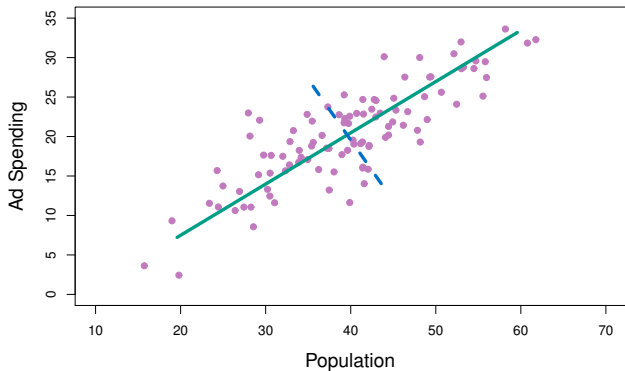
1. 주성분분석
2. 주성분회귀분석

주성분분석 I

주성분분석(principal component analysis)의 목적

1. p 개의 설명변수 X_1, X_2, \dots, X_p 가 있을 때, 설명변수들의 변동을 가장 크게 설명하는 새로운 저차원 변수들을 구하는 방법을 주성분분석이라 한다.
2. 새로운 변수들을 이용해 설명변수들 전체의 변동을 좀 더 쉽게 이해할 수 있다.

주성분분석 II



노트.

1. 2차원 자료가 그림에 주어져 있다. 이 2차원 자료의 변동을 가장 잘 설명하는 1차원 설명변수가 무엇일까?
2. 자료는 조그만 도시들의 인구수와 특정한 제품의 광고비로 이루어져 있다.

주성분분석 IV

주성분의 정의

1. 평균이 0인 설명변수들 X_1, X_2, \dots, X_p 의 첫번째 주성분은

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p, \quad \sum_{j=1}^p \phi_{j1}^2 = 1$$

와 같은 X_1, X_2, \dots, X_p 의 선형조합 중 분산이 가장 큰 선형조합이다.

2. $\phi_1 = (\phi_{11}, \phi_{21}, \dots, \phi_{p1})'$ 를 적재값(loadings)이라 한다.
3. 위의 그림의 경우

$$Z_1 = 0.839 \times (pop - \bar{pop}) + 0.544 \times (ad - \bar{ad})$$

가 첫번째 주성분이다. 평균이 0이 아닌 변수는 평균으로 빼준다.

주성분분석 V

4. 두번째 주성분은

$$Z_2 = \phi_{12}X_1 + \phi_{22}X_2 + \dots + \phi_{p2}X_p, \quad \sum_{j=1}^p \phi_{j2}^2 = 1, \phi_1 \perp \phi_2$$

를 만족하는 X_1, X_2, \dots, X_p 의 선형조합 중 분산이 가장 큰 선형조합이다. 여기서 $\phi_2 = (\phi_{12}, \phi_{22}, \dots, \phi_{p2})'$ 는 적재값이다.

5. 이와 같이, 세번째에서 p 번째 주성분 Z_3, \dots, Z_p 를 정의한다.

주성분분석 VI

미국의 강력범죄 자료

주어진 자료는 USArrest 자료로 미국 50개 주의 도시인구비율(UrbanPop), 폭행(Assault), 살인(Murder), 강간(Rape)의 인구 10만명당 체포수를 포함한 자료이다.

	PC1	PC2
Murder	0.5358995	-0.4181809
Assault	0.5831836	-0.1879856
UrbanPop	0.2781909	0.8728062
Rape	0.5434321	0.1673186

위의 표에서 첫번째 열의 벡터는 첫번째 주성분의 적재값, 두번째 열의 벡터는 두번째 주성분의 적재값이다. 첫번째 행의 벡터는 첫번째 변수 Murder의 주성분점수(principal component score), 두번째 행의 벡터는 두번째 변수 Assault의 주성분 점수이다.

노트. 주성분점수의 해석.

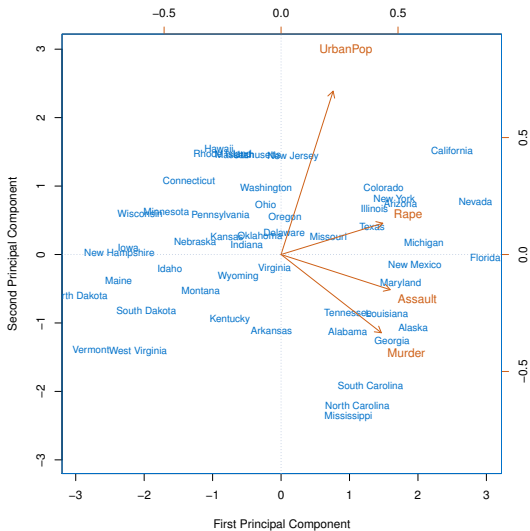
주성분분석 VII

주성분 점수로 각 변수가 각 주성분에 어떻게 영향을 미치는지 해석할 수 있다. 예를 들면 UrbanPop은 pc1에는 큰 영향을 미치지 않지만 pc2에는 많은 영향을 미친다.

pc1은 살인, 강도, 강간에 평균적으로 영향을 받고, pc2는 인구수에 주로 영향을 받는다. pc1은 범죄를, pc2는 인구수를 나타낸다고 할 수 있다. 즉, 살인, 강도, 강간의 체포수와 인구수로 이루어진 전체자료의 변동은 크게 범죄와 인구수의 변동으로 설명할 수 있다.

1. 주성분 분석의 어려움은 각 axis의 해석이 쉽지 않다는데 있다.
2. 변수의 centering을 하지 않은 경우는 첫 번째 축 평균성분 두번째 축은 자료를 clustering/partion 하는데 가장 영향력있는 선형 조합을 포신한다. 두 번째 축을 구성하는 방향 벡터를 fider 벡터라 이름짓고 특별하게 취급한다.

주성분분석 VIII



주성분분석 IX

노트.

첫번째 주성분과 두번째 주성분을 그린 그림이다.

노트. 그림에 대한 해석.

1. 그림의 파란 점은 각 관측치에 해당하는 주성분 즉

$$z_{1i} = \phi_{11}x_{1i} + \dots + \phi_{p1}x_{pi}$$

$$z_{2i} = \phi_{12}x_{1i} + \dots + \phi_{p2}x_{pi}$$

를 계산하여 (z_{1i}, z_{2i}) 를 그린 것이다. 즉, 각 관측치를 주성분의 좌표로 그린 것이다.

2. 그림의 오렌지 화살표는 각 변수에 해당하는 적재값을 우측과 위에 새로운 축을 그려서 그린 것이다. 예를 들면 x_1 에 해당하는 화살표는 (ϕ_{11}, ϕ_{12}) 를 그린 것이다. 이 값을 x_1 의 주성분점수(principal component scores)라 한다. 즉, 오렌지 화살표는 각 변수가 각주성분에 미치는 영향을 나타낸다. 즉, UrbanPop의 경우 x 축의 좌표는 0 근처, y 축의 좌표는 1 근처인데 이는 UrbanPop이 pc1에는 영향을 안 미치고, pc2에 주로 영향을 미친다는 것을 알 수 있다.

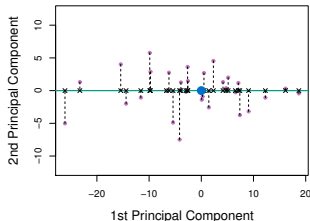
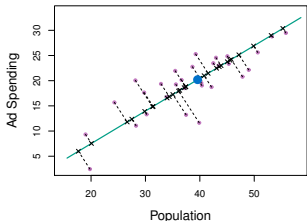
주성분분석 X

3. 이와 같은 그림을 **쌍도(biplot)라 한다**. 왜냐하면 이 그림이 주성분 적재값과 주성분점수(principal component scores)를 함께 그리기 때문이다.
4. 오렌지 화살표를 보면 변수 x_j 들이 주성분과 어떻게 관계되는지 알 수 있다. UrbanPop은 pc1에는 큰 영향을 미치지만 pc2에는 큰 영향을 미치지 않는다. Rape, Assault, Murder는 pc1에 비슷한 정도로 영향을 미친다. **또한 범죄 관련 변수들 Rape, Assault, Murder가 서로 가까이 위치하기 때문에 이들이 상관관계가 있다는 것을 알 수 있다. UrbanPop은 이들 범죄관련 변수들과 상관관계가 낮다.**
5. 파란점을 보면 각 주가 pc1과 pc2의 좌표로 어떻게 표시되는지 보인다. 예를 들면 Florida는 pc1은 크지만 pc2의 값은 거의 0에 가깝다. pc1은 주로 범죄관련 변수들로 설명되고 pc2는 주로 도시인구비율로 설명되므로, Florida의 전반적 범죄율이 높은 것을 알 수 있다. California, Nevada도 마찬가지이다. 반면에 North Dakota는 전반적 범죄율이 낮다.
PC score값을 통하여 주들의 군집분석이 가능하다.

주성분분석 XI

주성분분석의 또다른 해석

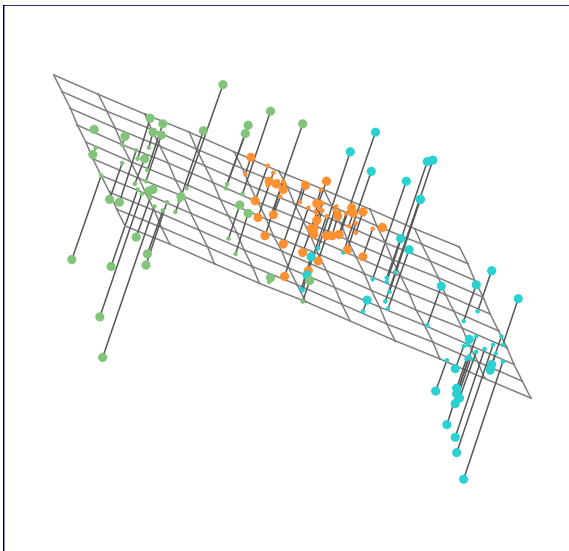
1. 첫번째 주성분은 X_1, \dots, X_p 의 공간에 하나의 벡터로 표현할 수 있다. 즉, 주성분 적재값은 변수들의 공간에서 하나의 벡터를 나타낸다. 모든 관측치를 이 벡터의 직선에 사영을 하고, 사영된 점과 원래 관측치 사이의 거리를 잔차라 하면, 첫번째 주성분은 이 잔차제곱합을 최소화 하는 벡터이다. 즉, 첫번째 적재값 벡터는 관측치들을 가장 잘 요약하는 벡터이다.



주성분분석 XII

2. 첫번째와 두번째 주성분은 위와 같이 두 적재값 벡터를 X_1, \dots, X_p 의 공간에 평면으로 나타낼 수 있다. 마찬가지로 관측치를 이 평면에 사영하고 그 차이를 잔차라 하면 이 평면은 잔차를 최소로 하는 평면이다. 즉, 첫번째와 두번째 적재값 벡터로 이루어지는 평면은 관측치들을 가장 잘 요약하는 평면이다.

주성분분석 XIII

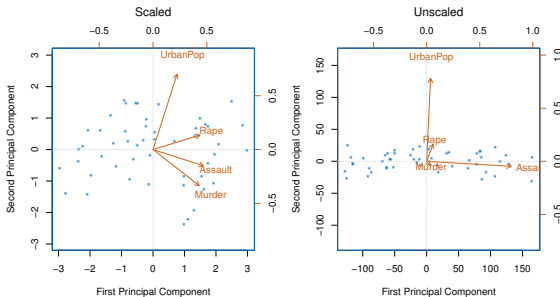


3. 3, 4, ... p 차원에 대해서도 동일한 해석을 할 수 있다.

주성분분석의 이슈들 I

척도화(scaling)

1. 변수들을 척도화 했을 때와 척도화하지 않았을 때, 주성분분석의 결과가 달라진다.



주성분분석의 이슈들 II

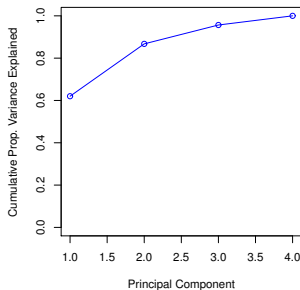
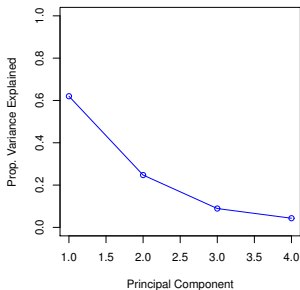
2. 왼쪽은 척도화했을 때, 오른쪽은 척도화하지 않을 때의 쌍도이다. Murder, Rape, Assault, UrbanPop의 분산은 각각 18.97, 87.73, 6943.16, 209.5이다. Assault의 분산이 가장 크기 때문에 첫번째 주성분은 Assault의 변동만 표현하게 된다. 오른쪽 그림이 이를 나타낸다.
3. 주성분분석하기 전에 대개 변수들을 척도화하는 것을 추천한다. 모든 변수가 동일한 단위일 때는 척도화하지 않아도 된다.

주성분의 유일성

주성분은 적재벡터의 부호를 제외하면 유일하다. 부호가 달라도 동일한 직선을 나타낸다.

주성분분석의 이슈들 III

산비탈그림(scree plot, 스크리그림) : 설명된 분산 비율



왼쪽 그림은 전체 분산합 중에서 주성분의 분산의 비율을, 오른쪽 그림은 주성분 분산의 누적합을 그린 것이다.

노트.

주성분분석의 이슈들 IV

1. 왼쪽 그림과 오른쪽 그림은 각각 다음을 그린다.

$$\frac{\mathbb{V}ar(Z_j)}{\sum_{j=1}^p \mathbb{V}ar(Z_j)} \text{와} \frac{\sum_{j=1}^k \mathbb{V}ar(Z_j)}{\sum_{j=1}^p \mathbb{V}ar(Z_j)}.$$

2. $\sum_{j=1}^p \mathbb{V}ar(Z_j) = \sum_{j=1}^p \mathbb{V}ar(X_j)$ 이다.

주성분분석의 이슈들 V

몇 개의 주성분을 사용하나?

산비탈그림을 보고 판단한다. 감소하는 추세가 느껴지는 부분에서 끊는다.

노트.

위의 그림은 2개의 주성분을 선택하는 것이 좋을 것 같다. 세번째 주성분의 분산은 전체의 10% 미만을 설명한다.

scree plot과 함께 $pr.var < 1$ /누적 설명력 80-90%도 자주 사용된다.

주성분분석 R 코드 I

```
str(USArrests)

## 'data.frame': 50 obs. of 4 variables:
## $ Murder : num 13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
## $ Assault : int 236 263 294 190 276 204 110 238 335 211 ...
## $ UrbanPop: int 58 48 80 50 91 78 77 72 80 60 ...
## $ Rape : num 21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...

head(USArrests)

##           Murder Assault UrbanPop Rape
## Alabama      13.2      236       58 21.2
## Alaska       10.0      263       48 44.5
## Arizona        8.1      294       80 31.0
## Arkansas       8.8      190       50 19.5
## California     9.0      276       91 40.6
## Colorado       7.9      204       78 38.7
```

미국 각 주의 범죄관련 자료이다. 50개의 관측치와 4개의 변수로 구성되어 있다.

주성분분석 R 코드 II

```
apply(USArrests, 2, mean)
```

```
##      Murder      Assault UrbanPop      Rape  
##      7.788    170.760    65.540    21.232
```

```
apply(USArrests, 2, sd)
```

```
##      Murder      Assault UrbanPop      Rape  
## 4.355510 83.337661 14.474763 9.366385
```

```
summary(USArrests)
```

```
##      Murder      Assault      UrbanPop      Rape  
## Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30  
## 1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07  
## Median : 7.250   Median :159.0   Median :66.00   Median :20.10  
## Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23  
## 3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18  
## Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
```

각 변수들의 평균과 분산은 매우 다르다. 주성분분석을 할 때, 척도화(scaling)를 하는 것이 필요하다.

주성분분석 R 코드 III

```
pr.out=prcomp(USArrests, scale=TRUE)
```

```
str(pr.out) 또는 head(pr.out)
```

```
## List of 5
## $ sdev      : num [1:4] 1.575 0.995 0.597 0.416
## $ rotation: num [1:4, 1:4] -0.536 -0.583 -0.278 -0.543 0.418 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:4] "Murder" "Assault" "UrbanPop" "Rape"
## .. ..$ : chr [1:4] "PC1" "PC2" "PC3" "PC4"
## $ center   : Named num [1:4] 7.79 170.76 65.54 21.23
## .. attr(*, "names")= chr [1:4] "Murder" "Assault" "UrbanPop" "Rape"
## $ scale    : Named num [1:4] 4.36 83.34 14.47 9.37
## .. attr(*, "names")= chr [1:4] "Murder" "Assault" "UrbanPop" "Rape"
## $ x        : num [1:50, 1:4] -0.976 -1.931 -1.745 0.14 -2.499 ...
## .. attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
## .. ..$ : chr [1:4] "PC1" "PC2" "PC3" "PC4"
## - attr(*, "class")= chr "prcomp"
```

척도화(scale=TRUE)를 해서 주성분분석을 하였다. 주성분분석의 결과는 5개의 리스트(sdev, rotation, center, scale, x)로 이루어져있다.

노트.

결과의 각 값들은 아래에 설명한다.

주성분분석 R 코드 IV

```
pr.out$center
```

```
##      Murder  Assault UrbanPop      Rape  
##      7.788   170.760   65.540    21.232
```

```
pr.out$scale
```

```
##      Murder    Assault  UrbanPop      Rape  
##      4.355510  83.337661 14.474763  9.366385
```

5개의 리스트 중 center와 scale은 주성분분석을 수행하기 전 각 변수들의 평균과 표준편차를 의미한다.

주성분분석 R 코드 V

pr.out\$rotation PC1의 loading값의 부호가 반대로 되어 있는데 이는 중요하지 않다.
축의 방향은 동일.

##		PC1	PC2	PC3	PC4
## Murder	-0.5358995	0.4181809	-0.3412327	0.64922780	
## Assault	-0.5831836	0.1879856	-0.2681484	-0.74340748	
## UrbanPop	-0.2781909	-0.8728062	-0.3780158	0.13387773	
## Rape	-0.5434321	-0.1673186	0.8177779	0.08902432	

rotation은 적재 벡터를 포함하고 있다. 즉,

$$\begin{aligned}pc_1 &= -0.535 \times Murder - 0.583 \times Assault \\&\quad -0.278 \times UrbanPop - 0.543 \times Rape \\pc_2 &= 0.418 \times Murder + 0.187 \times Assault \\&\quad -0.872 \times UrbanPop - 0.167 \times Rape\end{aligned}$$

와 같이 나타낼 수 있다.

주성분분석 R 코드 VI

```
dim(pr.out$x)

## [1] 50  4

str(pr.out$x)

##  num [1:50, 1:4] -0.976 -1.931 -1.745 0.14 -2.499 ...
##  - attr(*, "dimnames")=List of 2
##    ..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
##    ..$ : chr [1:4] "PC1" "PC2" "PC3" "PC4"
```

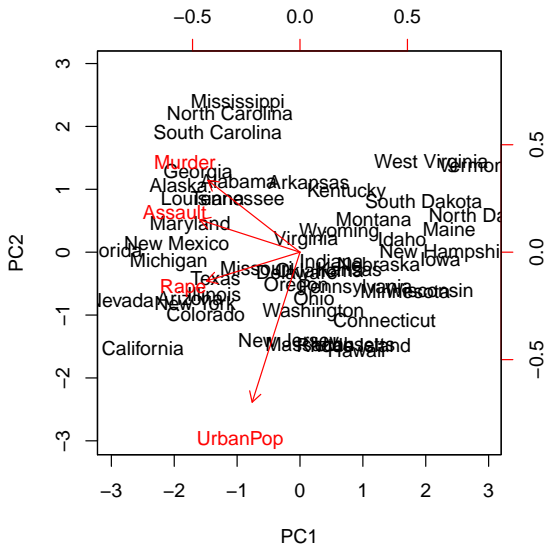
x는 주성분을 포함하고 있다. 즉, x[,1]은 첫번째 주성분, x[,2]는 두번째 주성분이다. 각 관측치마다 주성분의 값을 계산하여서 x의 행은 관측치의 개수와 같고 열은 주성분의 개수와 같다.

주성분분석 R 코드 VII

```
biplot(pr.out, scale=0)
```

cex와 같은 그림 option을 사용할 수 있다.

주성분분석 R 코드 VIII



주성분분석 R 코드 IX

쌍도를 그린 그림이다.

노트.

이 그림은

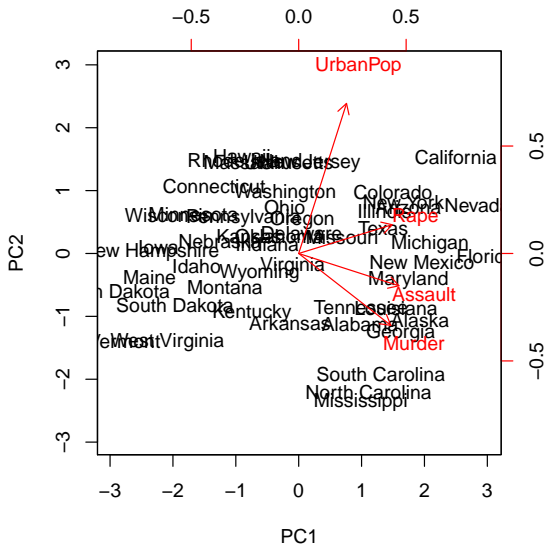
```
plot(pr.out$x[,1], pr.out$x[,2])
```

를 그린 것과 같다. 두번째와 세번째 주성분의 그림을 그리고 싶으면 옵션 `choices = c(2,3)`을 쓰면 된다. 변수는 λ^{scale} 와 같이 관측치는 $\lambda^{1-scale}$ 로 표시된다. **scale=0는 있는 그대로 척도를 맞추는 것이다.** λ 는 주성분분석의 특이값(singular value)를 의미한다.

주성분분석 R 코드 X

```
pr.out$rotation=-pr.out$rotation  
pr.out$x=-pr.out$x  
biplot(pr.out, scale=0)
```

주성분분석 R 코드 XI



주성분분석 R 코드 XII

주성분의 사인을 바꾼것이다.

```
pr.out$sdev

## [1] 1.5748783 0.9948694 0.5971291 0.4164494

pr.var=pr.out$sdev^2
pr.var

## [1] 2.4802416 0.9897652 0.3565632 0.1734301

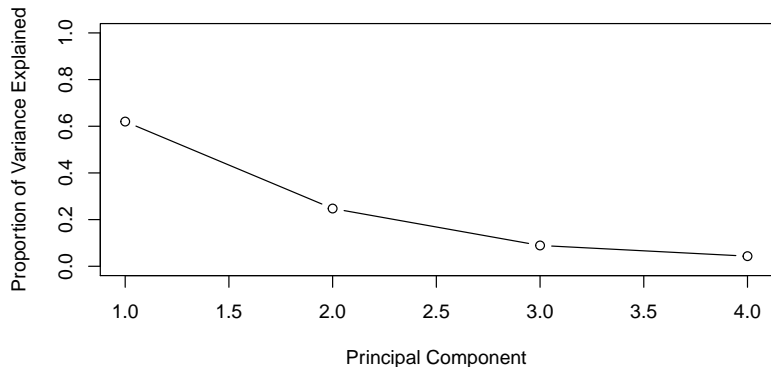
pve=pr.var/sum(pr.var) scree plot과 함께 pr.var<1/누적 설명력 80-90%도 자주 사용된다.
pve

## [1] 0.62006039 0.24744129 0.08914080 0.04335752
```

sdev는 주성분의 표준편차이다. 주성분의 분산을 계산하였다. 각 주성분의 설명 비율이다.

```
plot(pve, xlab="Principal Component", ylab="Proportion of Variance Explained", ylim=c(0,1))
```

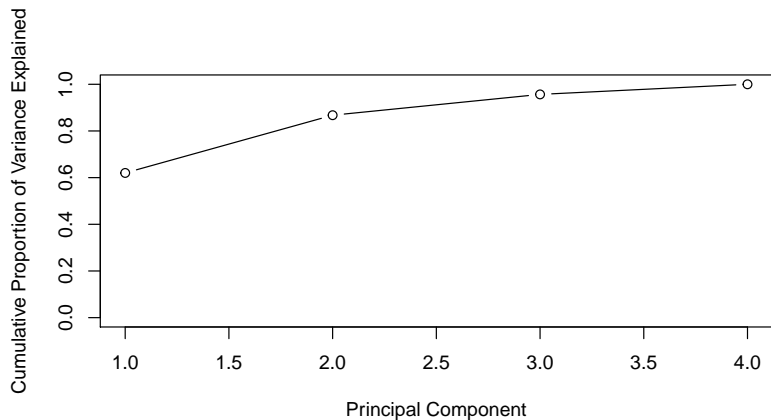
주성분분석 R 코드 XIII



각 주성분의 설명하는 분산 비율의 그림이다.

```
plot(cumsum(pve), xlab="Principal Component", ylab="Cumulative Proportion of Variance Explained")
```


주성분분석 R 코드 XIV



주성분의 누적 분산 비율이다.

주성분회귀분석(principal component regression) I

목적

1. Y 를 반응변수로 X_1, \dots, X_p 를 설명변수로 회귀모형을 적합하고자 한다. 이 때, 설명변수의 개수 p 가 클 때, 변수의 개수를 줄이는 방법으로 사용한다.
2. 설명변수들 사이에 공선형성(collinearity, 공선성)이 있을 때 공선형성을 없애는 방법으로 사용되기도 한다.

방법

1. 설명변수 X_1, \dots, X_p 에 주성분분석을 적용한 후 $M \leq p$ 개의 주성분 Z_1, \dots, Z_M 을 새로운 설명변수로 선택한다.
2. 선택된 M 개의 주성분 Z_1, \dots, Z_M 를 설명변수로 Y 를 반응변수로 회귀모형을 적합한다.

주성분회귀분석(principal component regression) II

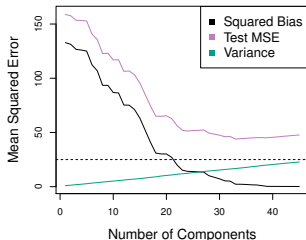
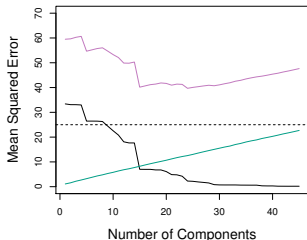
가정

1. 주성분회귀분석은 설명변수들의 분산이 큰 방향이 반응변수와 상관이 높은 방향이라는 가정하에서 사용된다.
2. 이 가정이 항상 옳은 것은 아니지만 종종 성립한다.

주성분 개수의 선택

1. 주성분의 개수 M 이 커질수록 주성분회귀모형의 유연성이 커진다. 따라서 M 이 커질수록 분산은 커지고 편이는 작아진다.
2. M 은 교차검증을 통해 선택한다.

주성분회귀분석(principal component regression) III



주성분 회귀분석의 성능

1. 설명변수의 변동을 설명하기 위해 많은 주성분이 필요하면, 능선회귀, 라쏘 보다 성능이 좋지 않고, 적은 수의 주성분이 필요하면 다른 방법들보다 성능이 좋다.

주성분회귀분석 R 코드 I

```
library(pls)

##
## Attaching package: 'pls'
##
## The following object is masked from 'package:stats':
##
##   loadings

library(ISLR)
```

노트.

Partial Least Squares Regression (PLSR), Principal Component Regression (PCR) and Canonical Powered Partial Least Squares (CPPLS)가 있는 패키지이다.

```
str(Hitters)
head(Hitters)
```

주성분회귀분석 R 코드 II

노트.

Hitters는 타자들의 연봉에 대한 자료로 322명의 타자들에 대한 20개의 변수를 포함하였다.

```
set.seed(2)
pcr.fit=pcr(Salary~., data=Hitters,scale=TRUE,validation="CV")
```

```
summary(pcr.fit)
```

```
> summary(pcr.fit)
Data:          X dimension: 263 19
Y dimension: 263 1
Fit method: svdpc
Number of components considered: 19
```

VALIDATION: RMSEP

Cross-validated using 10 random segments.

(Intercept)	1 comps	2 comps	3 comps	4 comps
-------------	---------	---------	---------	---------

주성분회귀분석 R 코드 III

CV	452	348.9	352.2	353.5	352.8
adjCV	452	348.7	351.8	352.9	352.1

....

TRAINING: % variance explained

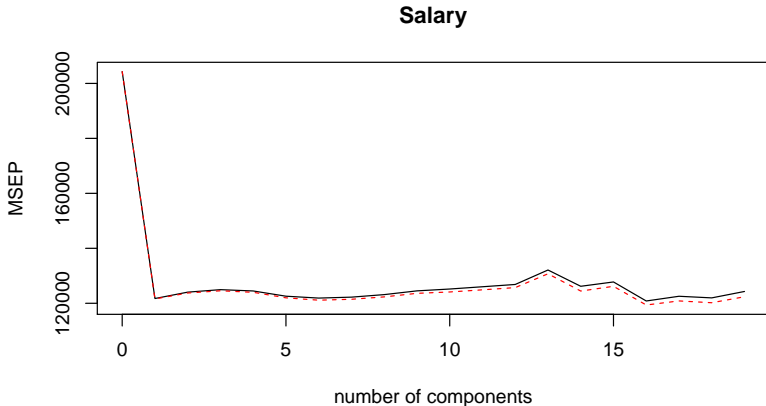
	1 comps	2 comps	3 comps	4 comps	5 comps
X	38.31	60.16	70.84	79.03	84.29
Salary	40.63	41.58	42.17	43.22	44.90

....

주성분회귀분석 R 코드 IV

함수 pcr의 사용법은 lm 함수와 비슷하다. scale은 설명변수를 척도화해서 주성분분석을 돌리는 것이다. validation="CV"는 주성분의 개수를 정하는데 10겹 교차검증을 사용하게 한다.

```
validationplot(pcr.fit, val.type="MSEP")
```



주성분회귀분석 R 코드 V

교차검증에러를 그림으로 그린 것이다. $M = 16$ 일 때 교차검증에러가 가장 작지만 $M = 1$ 일 때 이미 에러가 작다.

```
set.seed(1)
pcr.fit=pcr(Salary~., data=Hitters,subset=train,scale=TRUE, validation="CV")
validationplot(pcr.fit,val.type="MSEP")
pcr.pred=predict(pcr.fit,x[test,],ncomp=7)
mean((pcr.pred-y.test)^2)
pcr.fit=pcr(y~x,scale=TRUE,ncomp=7)
summary(pcr.fit)
```

훈련자료에 주성분회귀분석을 적합하고 시험오차를 계산하였다.
예측값을 계산할 때는 predict 함수를 사용한다.

```
Hitters=na.omit(Hitters)
x=model.matrix(Salary ~.,Hitters)[-1]
y=Hitters$Salary
set.seed(1)
train=sample(1:nrow(x), nrow(x)/2)
test=(-train)
y.test=y[test]
```

주성분 회귀의 회귀식 계산

USArrests 자료를 이용하고 가상 response에 대한 PCR을 생각하여 보자.

```
x5=x2+rnorm(50,0,1)
```

```
y=1+1*x1+2*x2+3*x3+4*x4+5*x5 [1,1,2,3,4,5]=참값
```

prcomp은 자료가 numeric 이어야 한다.

```
x0=USArrests
```

```
x5=USArrests[,2]+rnorm(50,0.01)
```

```
x=cbind(x0,x5)
```

```
x=scale(x)
```

```
y=1+1*x[1]+2*x[2]+3*x[3]+4*x[4]+5*x[5]+rnorm(50)
```

```
pr.out=prcomp(x,scale=T)
```

```
head(pr.out)
```

```
pcscore=pr.out$x[,1:2] #2개의 principal component 사용
```

```
load=pr.out$rotation[,1:2]
```

```
plm=lm(unlist(y)...pcscore)
```

```
pbeta=plm$coef
```

```
betaest=rep(0,6)
```

```
betaest[1]=pbeta[1]
```

```
for(k in (2:6)) {
```

```
    betaest[k]=pbeta[2]*load[k-1,1]+pbeta[3]*load[k-1,2]
```

```
}
```

```
> betaest
```

```
[1] 5.71191806 -0.06797535 -0.01945628 0.20801016  
0.06630954 -0.01876980
```

```
>
```

```
> a=lm(unlist(y)~x)
```

```
> summary(a)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.7119	0.1632	34.994	<2e-16 ***
xMurder	0.2511	0.2880	0.872	0.3879
xAssault	2.1448	13.6173	0.158	0.8756
xUrbanPop	0.3890	0.1980	1.965	0.0557 .
xRape	-0.1911	0.2422	-0.789	0.4343
xx5	-2.3211	13.5938	-0.171	0.8652

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.154 on 44 degrees of freedom

참고문헌

아래의 책에서 제공하는 그림들을 사용하였다.

1. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Springer, 2013.