- Data Mining Principles
- MSc in Analytics

- Dr. Anil D Chaturvedi
- Email: anilchaturvedi@uchicago.edu
- Phone: 301-299-2434

THE UNIVERSITY OF CHICAGO

- Good Morning.
- Welcome to Data Mining Principles
- Session 4

Data Mining Principles
Session 4: Dimension Reduction

1. Clarifications on Assignment 1
2. Principal Components based Factor Analysis
3. General Factor Analysis
4. Sneak Peak at 3-way Hybrid Factor Analysis
5. Conclusion

Clarifications from Assignment 1

1. Excellent performance overall.

2. SCALE DATA! It is critical. **Normalize. Normalize. Normalize.**

3. Hold-outs: Not 1:500 and 500:1000 Data could be sorted. Not random. Please take random samples.

4. Assume you selected 3 clusters in training sample. Use centers of THIS solution as starting point for a 3-cluster solution in Holdout. You run K-means once only in Holdout. Many of you ran 2…K in Holdout too. That is ok. But it is a LOT of work.

5. Don't compare times of komeans and kmeans. Unfair comparison. Komeans was written in R for YOU. Kmeans is written in c/c++. Komeans in current form in R is bound to be slower.

6. Use means of clusters to INTERPRET clusters. Some of you did. Many did not.

Clarifications for Assignments

1. Use function komeans(). Other functions are called by komeans.
2. Don't have to use all clusters! Only some will be big enough and meaningful. The others might be too small or unimportant. But in the presence of many smaller ones, the bigger clusters get tighter, and are more interpretable! They are Partitions.
3. Learn function apply, lapply, tapply, xtabs. It makes generating Cluster profiles lot easier.
4. Learn subset()
5. Go to R training

Data Mining Principles
Session 4: Dimension Reduction

Factor Analysis Based on Principal Components
or
Principal Components Analysis

Bilinear models

$$d_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr}$$

$$\mathrm{D}_{IxJ} = \mathrm{A}_{IxR} \mathrm{B}'_{RxJ}$$

Examples of Bilinear models: Covariance

$$\operatorname{cov}(x, y) = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$s = \frac{1}{n-1}\mathbf{x}'\mathbf{y} - \frac{1}{n(n-1)}(\mathbf{x}'\mathbf{1})(\mathbf{y}'\mathbf{1})$$

If x and y are centered, then

$$s = \frac{1}{n-1}\mathbf{x}'\mathbf{y}$$

Examples of Bilinear models: Covariance

If x and y are centered, then

$$\mathbf{s} = \frac{1}{n-1}\mathbf{x}'\mathbf{y}$$

$$\mathrm{s} \propto \mathbf{x}'\mathbf{y}$$

If x and y are standardized, then

$$\rho = \frac{1}{n-1}\mathbf{x}'\mathbf{y}$$

$$\rho \propto \mathbf{x}'\mathbf{y}$$

Examples of Bilinear models: Covariance

If Data D containing n observations on p variables are centered, then

$$\mathbf{COV}_{\mathbf{p\,x\,p}} = \frac{1}{n-1}\mathbf{D}'_{\mathbf{p\,x\,n}}\mathbf{D}_{\mathbf{n\,x\,p}}$$

$$\mathbf{COV}_{\mathbf{p\,x\,p}} \propto \mathbf{D}'_{\mathbf{p\,x\,n}}\mathbf{D}_{\mathbf{n\,x\,p}}$$

If D centered and unit variance for each variable (standardized), then

$$\mathbf{COR}_{\mathbf{p\,x\,p}} = \frac{1}{n-1}\mathbf{D}'_{\mathbf{p\,x\,n}}\mathbf{D}_{\mathbf{n\,x\,p}}$$

$$\mathbf{COR}_{\mathbf{p\,x\,p}} \propto \mathbf{D}'_{\mathbf{p\,x\,n}}\mathbf{D}_{\mathbf{n\,x\,p}}$$

Examples of Bilinear models: Dimension Reduction via Factor Analysis

$$d_{ij} \approx \sum_{r=1}^{R} f_{ir} l_{jr}$$

$$\mathrm{D}_{IxJ} \approx \mathrm{F}_{IxR} \mathrm{L}_{RxJ}$$

- In Factor Analysis only **numeric** Data D_{ij} are known
- Factor Scores F are unknown
- Factor Loadings L are unknown
- Unique variance

Examples of Bilinear models: Column Reduction via Principal Components

$$d_{ij} \approx \sum_{r=1}^{R} f_{ir} l_{jr}$$

$$\mathrm{D}_{IxJ} \approx \mathrm{F}_{IxR} \mathrm{L}_{RxJ}$$

- In Factor Analysis only numeric Data D_{ij} are known
- Factor Scores F are unknown
- Factor Loadings L are unknown
- F and L are orthogonal

Examples of Multilinear models: Projection to a vector

- Post multiplying Data **D** of k variables by a vector **c of unit length** is equivalent to projection of each row (vector) of Data on vector **c**

$$\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k} \\ d_{21} & d_{22} & \dots & d_{2k} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nk} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_k \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{bmatrix}$$

$$D_{nxk} V_{kx1} = p_{nx1}$$

$$\begin{bmatrix} \mathbf{d}'_1 \\ \mathbf{d}'_2 \\ \dots \\ \mathbf{d}'_n \end{bmatrix} \mathbf{c} = \mathbf{p} \quad \Rightarrow \quad \begin{matrix} \mathbf{d}'_1\mathbf{c} = \mathrm{p}_1 \\ \mathbf{d}'_2\mathbf{c} = \mathrm{p}_2 \\ \dots \\ \mathbf{d}'_n\mathbf{c} = \mathrm{p}_n \end{matrix}$$

Examples of Bilinear models: Orthogonal vectors Bases of Support Vectors

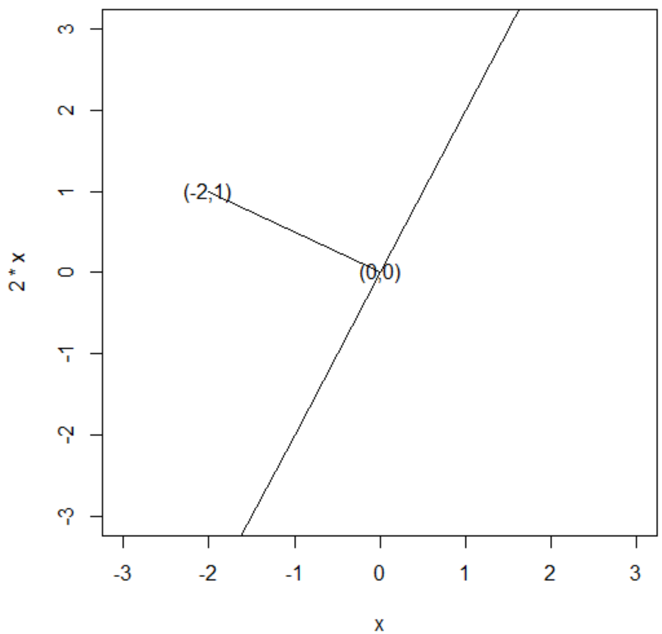
- Let y = 2x be a two dimensional line
- Then (x,y) = (0,0) and (1,2) are points of line
- What is the Equation of line orthogonal to it?
- **Rewrite equation as -2x +y = 0**
- **(-2 1) (x y)^T = 0**
- **The row vector of coefficients (-2 1) is ORTHOGONAL to points on line y = 2x**

$$y = 2x \Rightarrow$$

$$-2x + y = 0 \Rightarrow$$

$$\begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$c'x = 0 \Rightarrow$ the vectors c and x are orthogonal



- **x=seq(-3,3)**
- **plot(x, 2*x,type="l",xlim=c(-3,3),ylim=c(-3,3))**
- **lines(c(0,-2), c(0,1))**
- **text(c(0,-2), c(0,1),labels = c("(0,0)","(-2,1)"))**

Why Principal Component Analysis

1. Large number of numeric predictor variables. Say 1000’s
   - Pair-wise correlations – too many
   - Handle to summarize data

2. Pattern Detection
   - Which are similar and which are dissimilar?

Why Principal Component Analysis

3. To minimize the association among the variables
   - Too many variables that are highly correlated or measuring the same concept or construct

4. Discover the underlying theme or construct being measured by a set of variables e. g.,
   - Quantitative Aptitude
   - Verbal Aptitude
   - Logical Aptitude

Principal Component Analysis: Illustration

- Assume we have n x p numeric data matrix D
  - n is number of rows of observations
  - p is number of columns

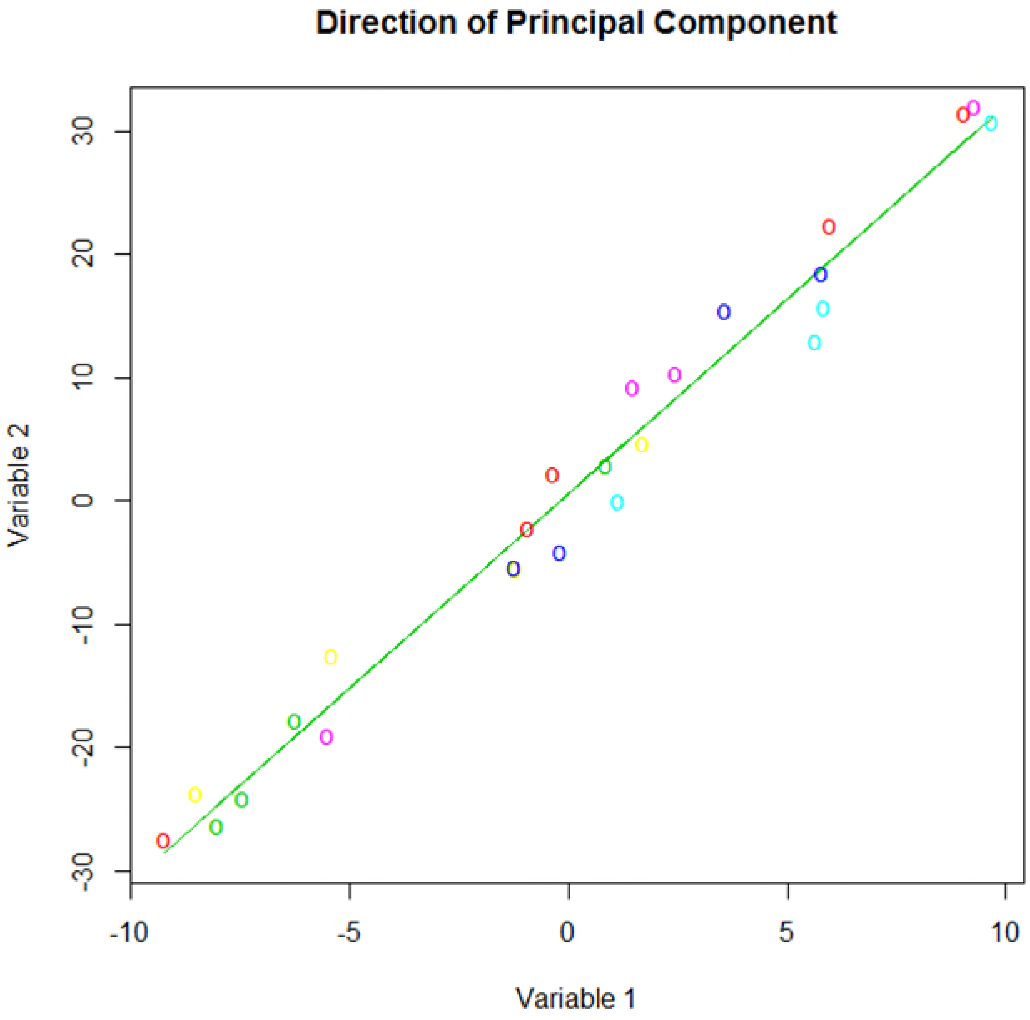
- Assume that matrix is **standardized** (i.e. each column has mean = 0 and variance = 1)

- **What is the direction in the p-dimensional space in which the data has maximal variance?**

Principal Component Analysis: Illustration

Number	x_1	x_2
1	-0.4	2.3
2	-8.0	-26.3
3	3.6	15.5
4	5.8	15.7
5	2.4	10.4
6	-1.2	-5.4
7	-0.9	-2.1
8	-7.4	-24.1
9	-1.2	-5.4
10	5.6	13.0
11	9.3	32.1
12	1.7	4.7
13	-9.3	-27.4
14	0.8	2.9
15	-0.2	-4.1
16	1.2	0.1
17	1.5	9.3
18	-8.5	-23.8
19	6.0	22.4
20	-6.3	-17.8
21	5.8	18.6
22	9.7	30.9
23	-5.5	-19.0
24	-5.4	-12.5
25	9.0	31.5

Principal Component Analysis: Illustration



Principal Component Analysis: The Principle

What is the direction in the p-dimensional space in which the data D has maximal variance?
Projections in a direction c are given by:

$$\mathbf{z}_{Ix1} = \mathbf{D}_{IxP}\mathbf{c}_{Px1}$$

Principal Component Analysis: The Principle

What is the direction in the p-dimensional space in which the data D has maximal variance? Projections in a direction c are given by:

$$\mathbf{z}_{Ix1} = \mathbf{D}_{IxP}\mathbf{c}_{Px1}$$

We can maximize variance by arbitrarily choosing elements of vector **c** arbitrarily large.

Principal Component Analysis: The Principle

What is the direction in the p-dimensional space in which the data D has maximal variance?

Projections in a direction c are given by:

$$\mathbf{z}_{Ix1} = \mathbf{D}_{IxP}\mathbf{c}_{Px1}$$

We can maximize variance by arbitrarily choosing elements of vector **c** arbitrarily large.

So we constrain the length of c to be = 1. that is

$$\mathbf{c}'_{1xP}\mathbf{c}_{Px1} = 1$$

Principal Component Analysis: The Principle

The variance of new points **z** in the direction of **c** **is given by**

$$\frac{1}{n-1}\mathbf{z}'\mathbf{z} = \frac{1}{n-1}(\mathbf{Dc})'(\mathbf{Dc})$$

$$= \frac{1}{n-1}\mathbf{c}'\mathbf{D}'\mathbf{D}\mathbf{c} = \mathbf{c}'\left(\frac{1}{n-1}\mathbf{D}'\mathbf{D}\right)\mathbf{c} = \mathbf{c}'\mathbf{R}\mathbf{c}$$

Where **R** is the correlation matrix of the input variables

Principal Component Analysis: The Principle

Choose **c** to maximize

$$\mathbf{c}'\mathbf{R}\mathbf{c}$$

Subject to

$$\mathbf{c}'\mathbf{c} = 1$$

Principal Component Analysis: The Principle

This is a Non-Linear optimization problem.
Solved using Lagrangians.

$$\text{Maximize } L = \mathbf{c}'\mathbf{R}\mathbf{c} - \lambda(\mathbf{c}'\mathbf{c} - 1)$$

Where λ is called the Lagrangian multiplier

Principal Component Analysis: The Principle

Solved using Lagrangian. Maximize

$$\text{Maximize } L = c'Rc - \lambda(c'c - 1)$$

Taking Derivatives with respect to elements of **c** yields:

$$\text{Maximize } \frac{\partial L}{\partial \mathbf{c}} = 2\mathbf{Rc} - 2\lambda\mathbf{c}$$

Principal Component Analysis: The Principle

Taking Derivatives with respect to elements of **c** yields:

$$\frac{\partial L}{\partial \mathbf{c}} = 2\mathbf{Rc} - 2\lambda\mathbf{c} = \mathbf{0}$$

$$\Rightarrow \mathbf{Rc} = \lambda\mathbf{c} = \lambda\mathbf{Ic}$$

Principal Component Analysis: The Principle

$$\Rightarrow \mathbf{Rc} = \lambda\mathbf{c} = \lambda\mathbf{Ic}$$

$$\Rightarrow (\mathbf{R} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$$

This is called the Characteristic Equation
Also the Eigen-Value Eigen-Vector
Problem. What does this Equation tell us?

Principal Component Analysis: The Principle

$$\Rightarrow \mathbf{Rc} = \lambda\mathbf{c} = \lambda\mathbf{Ic}$$

$$\Rightarrow (\mathbf{R} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$$

This is called the Characteristic Equation
Also the Eigen-Value Eigen-Vector Problem

Principal Component Analysis: The Principle

$$\Rightarrow \mathbf{Rc} = \lambda\mathbf{c} = \lambda\mathbf{Ic}$$

$$\Rightarrow (\mathbf{R} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$$

If R is full-rank (no fully collinear variables or no additive dependencies), then with p variables there will be p positive Eigenvalues.

Principal Component Analysis: The Principle

$$\Rightarrow \operatorname{var}(z) = c'Rc = c'\lambda c = \lambda c'c = \lambda$$

The Eigen value is the variance of the first Principal Component (Scores) or projections of the Data.

Each principal component is orthogonal to the others. Correlation between principal components is 0.

Principal Component Analysis: The Principle

If we extract all P principal components, then

$$\mathbf{z}_{Ix1} = \mathbf{D}_{IxP}\mathbf{c}_{Px1} \qquad \mathbf{c}'_{1xP}\mathbf{c}_{Px1} = 1$$

becomes

$$\mathbf{Z}_{IxP} = \mathbf{D}_{IxP}\mathbf{C}_{PxP} \qquad \mathbf{C}'_{\mathrm{PxP}}\mathbf{C}_{\mathrm{PxP}} = \mathbf{I}$$

Each principal component is orthogonal to the others. Correlation between principal components is 0.

Principal Component Analysis: Many Principal components (p)

$$\Rightarrow \mathrm{var}(z) = c'Rc = c'\lambda c = \lambda c'c = \lambda$$

$$\mathrm{var}(Z) = \mathrm{tr}(C'RC) = \mathrm{tr}(RC'C) = \mathrm{tr}(RI) = \mathrm{tr}(R)$$

$$C'C = I$$

Each principal component is orthogonal to the others. Correlation between principal components is 0.

$\mathrm{var}(c_i) = 1$, and $\mathrm{cov}(c_i, c_j) = 0$

$\mathrm{var}(z_i) = \lambda_i$, and $\mathrm{cov}(z_i, z_j) = 0$

Principal Component Analysis: Many Principal components (p)

Each principal component is orthogonal to the others. Correlation between principal components is 0.

Component Scores

$var(z_i) = \lambda_i$, and $cov(z_i, z_j) = 0$

Component Loadings

$var(c_i) = 1$, and $cov(c_i, c_j) = 0$

Principal Component Analysis: The Principle

Finally,

$$\mathbf{Z}_{IxP} = \mathbf{D}_{IxP}\mathbf{C}_{PxP} \qquad \mathbf{C}'_{PxP}\mathbf{C}_{PxP} = \mathbf{I}_{PxP}$$

Can be written as

$$\mathbf{Z}_{IxP}\mathbf{C}'_{PxP} = \mathbf{D}_{IxP}\mathbf{C}_{PxP}\mathbf{C}'_{PxP} = \mathbf{D}_{IxP}\mathbf{I}_{PxP} = \mathbf{D}_{IxP}$$

Or equivalently

$$\mathbf{D}_{IxP} = \mathbf{Z}_{IxP}\mathbf{C}'_{PxP}$$

Principal Component Analysis. Can be Solved via Singular Value Decomposition (SVD) of D

$$\mathbf{D}_{\mathrm{IxP}} = \mathbf{Z}_{\mathrm{IxP}} \mathbf{C}'_{\mathrm{PxP}}$$

$$\mathbf{D}_{\mathrm{IxP}} = \mathbf{P}_{\mathrm{IxP}} \mathbf{\Delta}_{\mathrm{PxP}} \mathbf{Q}'_{\mathrm{PxP}}$$

Principal Component Analysis. Can also be Solved via Eigen Vector Decomposition of Symmetric R

$$\mathbf{D}_{IxP} = \mathbf{Z}_{IxP}\mathbf{C}'_{PxP}$$

$$\mathrm{var}(\mathbf{D}) = \mathrm{var}(\mathbf{ZC}')$$

$$\frac{1}{n-1}\mathbf{D}'\mathbf{D} = \frac{1}{n-1}(\mathbf{ZC}')'(\mathbf{ZC}')$$

$$\mathbf{R} = \mathbf{C}\left(\frac{1}{n-1}\mathbf{Z}'\mathbf{Z}\right)\mathbf{C}'$$

$$\mathbf{R} = \mathbf{C}\Lambda\mathbf{C}'$$

Principal Component Analysis
Interpretation of Components

$$\mathbf{D}_{IxP} = \mathbf{Z}_{IxP}\mathbf{C}'_{PxP}$$

$$d_{ip} = \sum_{k=1}^{K} z_{ik} c_{pk}$$

$$d_{ip} = z_{i1}c_{p1} + z_{i2}c_{p2}$$

$$\mathbf{d}_{.p} = \mathbf{z}_{.1}\mathbf{c}_{p1} + \mathbf{z}_{.2}\mathbf{c}_{p2}$$

Principal Component Analysis Interpretation of Loadings as Correlations between data and components: Illustration

$$\begin{aligned}
\operatorname{cov}(d_{.p}, z_{.1}) &= \operatorname{cov}(z_{.1}c_{p1} + z_{.2}c_{p2}, z_{.1}) \\
&= \operatorname{cov}(z_{.1}c_{p1}, z_{.1}) + \operatorname{cov}(z_{.2}c_{p2}, z_{.1}) \\
&= c_{p1}\operatorname{cov}(z_{.1}, z_{.1}) + c_{p2}\operatorname{cov}(z_{.2}, z_{.1}) \\
&= c_{p1}\operatorname{cov}(z_{.1}, z_{.1}) + 0
\end{aligned}$$

$$\Rightarrow c_{p1} = \frac{\operatorname{cov}(d_{.p}, z_{.1})}{\operatorname{cov}(z_{.1}, z_{.1})} = \frac{\operatorname{cov}(d_{.p}, z_{.1})}{\operatorname{var}(z_{.1})} = \frac{\operatorname{cov}(d_{.p}, z_{.1})}{\operatorname{var}(z_{.1})\operatorname{sdev}(d_{.p})} = \frac{\rho_{p1}}{\operatorname{sdev}(z_{.1})}$$

$$\Rightarrow \rho_{p1} = \operatorname{sdev}(z_{.1})c_{p1}$$

Principal Component Analysis: Summary

1. Bilinear model
2. Based on Spectral Decomposition of Data D (or Eigen Vector Decomposition) or correlation matrix R
3. Seeks vectors (Loadings) of unit length (linear combinations of variables c) that result in projections (Component Scores) of data along direction having maximal variation
4. The vectors (loadings) are all mutually orthogonal and of unit length
5. The Component scores are orthogonal, and variance equals corresponding Eigen value

Principal Component Analysis: The Principle

Any Questions?

Principal Component Analysis: Loadings USArrests Data

head(USArrests)

50 Row for Each State, 4 Variables

	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

Principal Component Analysis: Attributes of PCA Results: USArrests Data

```
x=prcomp(USArrests,scale=TRUE)

> attributes(x)
$names
[1] "sdev"     "rotation" "center"   "scale"
"x"

$class
[1] "prcomp"
```

Principal Component Analysis: Loadings USArrests Data

```
x=prcomp(USArrests,scale=TRUE)
round(head(x$rotation),2)
```

	PC1	PC2	PC3	PC4
Murder	0.54	-0.42	0.34	-0.65
Assault	0.58	-0.19	0.27	0.74
UrbanPop	0.28	0.87	0.38	-0.13
Rape	0.54	0.17	-0.82	-0.09

Principal Component Analysis: Components USArrests Data

```
x=prcomp(USArrests,scale=TRUE)
attributes(x)
round(head(x$x),2)
```

	PC1	PC2	PC3	PC4
Alabama	0.98	-1.12	0.44	-0.15
Alaska	1.93	-1.06	-2.02	0.43
Arizona	1.75	0.74	-0.05	0.83
Arkansas	-0.14	-1.11	-0.11	0.18
California	2.50	1.53	-0.59	0.34
Colorado	1.50	0.98	-1.08	0.00

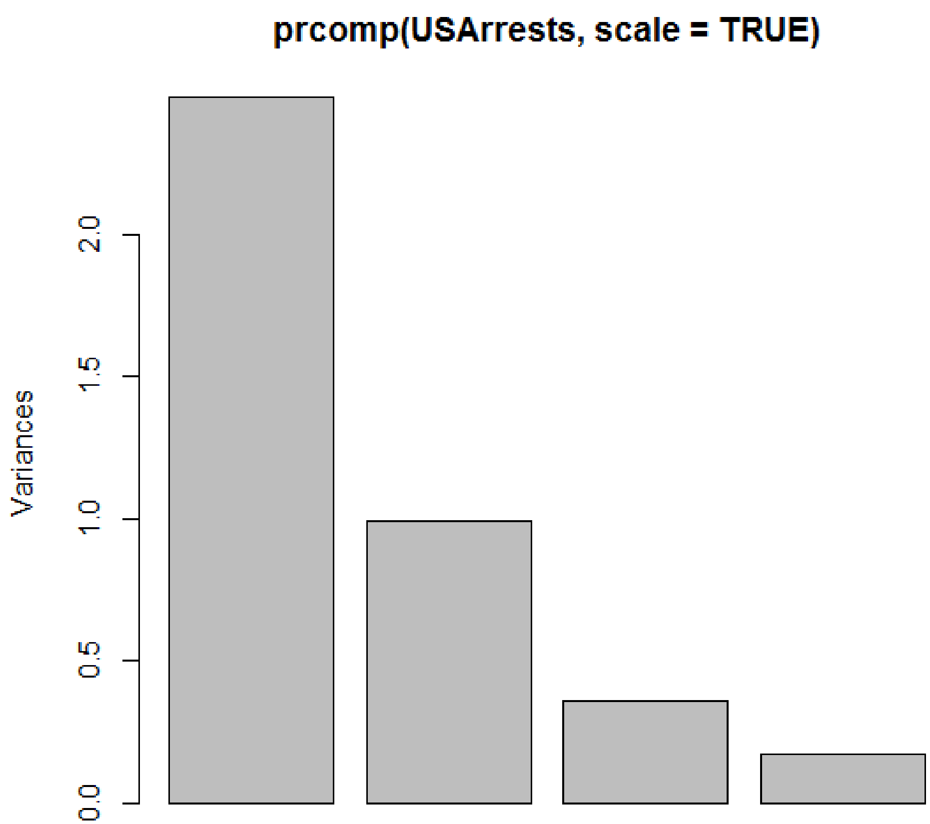
Principal Component Analysis: Loadings and Components: USArrests Data

	PC1	PC2	PC3	PC4
Alabama	0.98	-1.12	0.44	-0.15
Alaska	1.93	-1.06	-2.02	0.43
Arizona	1.75	0.74	-0.05	0.83
Arkansas	-0.14	-1.11	-0.11	0.18
California	2.50	1.53	-0.59	0.34
Colorado	1.50	0.98	-1.08	0.00

Loadings

Murder	0.54	-0.42	0.34	-0.65
Assault	0.58	-0.19	0.27	0.74
UrbanPop	0.28	0.87	0.38	-0.13
Rape	0.54	0.17	-0.82	-0.09

Principal Component Analysis: The Principle

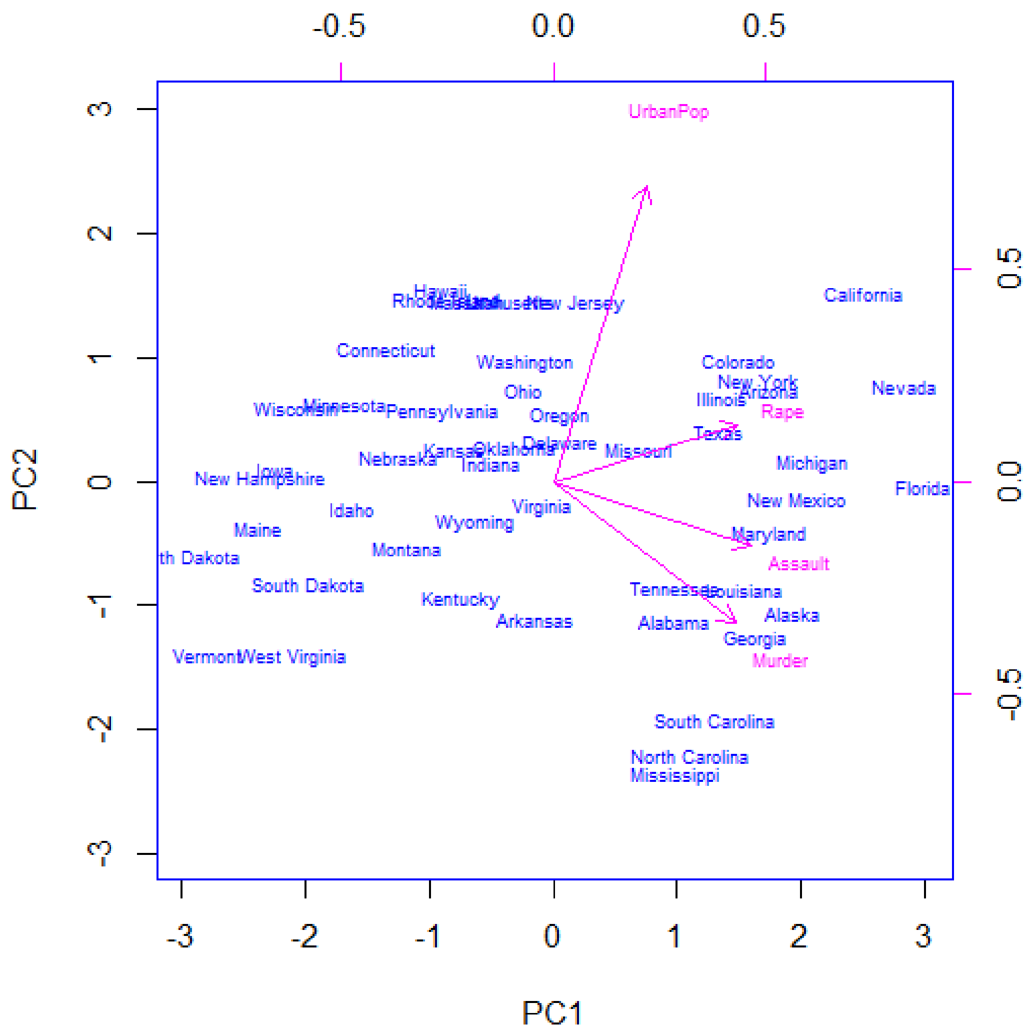


Principal Component Analysis: The Principle

Importance of components:

	PC1	PC2	PC3	PC4
Standard deviation	1.5749	0.9949	0.59713	0.41645
Proportion of Variance	0.6201	0.2474	0.08914	0.04336
Cumulative Proportion	0.6201	0.8675	0.95664	1.00000

Principal Components Analysis: Biplot



Principal Component Analysis: Illustration

```
x=prcomp(USArrests,scale=TRUE)
plot(x)
summary(x)
biplot(-x$x,-x$rotation,cex=0.6,col=c(4,6))

round(t(x$rotation) %*% x$rotation,2)
    PC1 PC2 PC3 PC4
PC1   1   0   0   0
PC2   0   1   0   0
PC3   0   0   1   0
PC4   0   0   0   1

round(x$rotation %*% t(x$rotation),2)
         Murder Assault UrbanPop Rape
Murder        1       0        0    0
Assault       0       1        0    0
UrbanPop      0       0        1    0
Rape          0       0        0    1
```

Principal Components Analysis: Illustration

```
x=prcomp(USArrests,scale=TRUE)
plot(x)
summary(x)
biplot(-x$x,-x$rotation,cex=0.6,col=c(4,6))

round(t(x$rotation) %*% x$rotation,2)
round(x$rotation %*% t(x$rotation),2)

round(cov(x$x),2)
round(t(x$x) %*% x$x/49,2)

sum(round(t(x$x) %*% x$x/49,2))
sum(x$sdev^2)

apply(x$rotation^2,1,sum)
apply(x$rotation^2,2,sum)

round(cor(x$x,USArrests),2)
round(diag(x$sdev) %*% t(x$rotation),2)
```

Principal Component Analysis: Illustration

```
x=prcomp(USArrests,scale=TRUE)
cor(as.vector(scale(USArrests)), as.vector(x$x[,1:2] %*% t(x$rotation)[1:2,]))^2
summary(x)

a = nrow(USArrests)
sa1=sample(1:a,as.integer(a*.7),replace=FALSE)
G=scale(USArrests)
G1= G[sa1,]
G2= G[-sa1,]
x=prcomp(G1,retx=TRUE)
round(cor(as.vector(G1), as.vector(x$x %*% t(x$rotation))),2)
y= predict(x, newdata=G2)
round(cor(as.vector(G2), as.vector(y %*%  t(x$rotation))),2)
plot(x)
 ###############################HOLDOUT WITH  FEWER COMPONENTS THAN 4 ###
a = nrow(USArrests)
sa1=sample(1:a,as.integer(a*.7),replace=FALSE)
x=prcomp(scale(USArrests[sa1,]),retx=TRUE,tol=0.4)
round(cor(as.vector(scale(USArrests[sa1,])), as.vector(x$x[,1:2] %*% t(x$rotation)[1:2,])),2)
y= predict(x, newdata=scale(USArrests[-sa1,]))
round(cor(as.vector(scale(USArrests[-sa1,])), as.vector(y[,1:2] %*%  t(x$rotation)[1:2,])),2)
```

Principal Components
Assumption of "common" principal components

$$\mathbf{D}_{\mathrm{IxP}} \approx \mathbf{Z}_{\mathrm{IxR}} \mathbf{C}'_{\mathrm{RxP}}$$

- In case we only extract R (< P) components, unexplained error left over.

- The PCA model explains variance in data as a function on the "common" principal components

Principal Component Analysis

Any Questions on PCA?

Factor Analysis

$$\mathbf{D}_{\mathrm{IxP}} \approx \mathbf{Z}_{\mathrm{IxR}} \mathbf{C}'_{\mathrm{RxP}}$$

$$\Rightarrow \mathrm{D}_{\mathrm{IxP}} = \mathrm{Z}_{\mathrm{IxR}} \mathrm{C}'_{\mathrm{RxP}} + \mathrm{E}_{\mathrm{IxP}}$$

Extract R (<P and < I) common factors

- Based on Maximum Likelihood
- Least Squares
- Weighted Least Squares
- Other approaches

Factor Analysis
The Common Factor Model with Error Variances

$$\mathbf{D}_{\mathrm{IxP}} \approx \mathbf{Z}_{\mathrm{IxR}}\mathbf{C}'_{\mathrm{RxP}}$$
$$\Rightarrow \mathrm{D}_{\mathrm{IxP}} = \mathrm{Z}_{\mathrm{IxR}}\mathrm{C}'_{\mathrm{RxP}} + \mathrm{E}_{\mathrm{P}}$$

- **Assumes Error Variances for each of the P variables that CANNOT be explained by the common factors.**

Factor Analysis

$$\operatorname{var}\left(d_{.p}\right)=\operatorname{var}\left(z_{.1}c_{p1}+z_{.2}c_{p2}+\ldots.\right)+\operatorname{var}\left(e_{p}\right)+2\operatorname{cov}\left(z_{.1}c_{p1}+z_{.2}c_{p2}+\ldots.,e_{p}\right)$$

$$\operatorname{var}\left(\mathbf{D}\right)=\operatorname{var}\left(\mathbf{ZC}'\right)+\operatorname{var}\left(\mathbf{E}\right)+0$$

$$\frac{1}{n-1}\mathbf{D}'\mathbf{D}=\frac{1}{n-1}\left(\mathbf{ZC}'\right)'\left(\mathbf{ZC}'\right)+\frac{1}{n-1}\mathbf{E}'\mathbf{E}$$

$$\mathbf{R}=\mathbf{C}\left(\frac{1}{n-1}\mathbf{Z}'\mathbf{Z}\right)\mathbf{C}'+\frac{1}{n-1}\mathbf{E}'\mathbf{E}$$

$$\mathbf{R}=\mathbf{CC}'+\boldsymbol{\triangleleft}$$

Principal Component Analysis
Illustration of Variance of a variable explained by 2 Common Factors

$$\operatorname{var}\left(d_{.p}\right)=\operatorname{var}\left(z_{.1}c_{p1}+z_{.2}c_{p2}+\ldots.\right)+\operatorname{var}\left(e_{p}\right)+2\operatorname{cov}\left(z_{.1}c_{p1}+z_{.2}c_{p2}+\ldots.,e_{p}\right)$$

$$\operatorname{var}\left(\mathbf{D}\right)=\operatorname{var}\left(\mathbf{ZC'}\right)+\operatorname{var}\left(\mathbf{E}\right)$$

$$\mathbf{R}=\mathbf{CC'}+\mathbf{\triangleleft}$$

- Δ is diagonal matrix of error variances
- Error variances for each of the P variables are typically estimated before the common factors are estimated
- Orthogonality of factor components is attempted but not guaranteed

Factor Analysis
Communality

$$\mathrm{var}\left(d_{.p}\right) \text{ explained by common factors: Communality}$$

$$= \mathrm{var}\left(z_{.1}c_{p1} + z_{.2}c_{p2}\right)$$

$$= \mathrm{var}\left(z_{.1}c_{p1}\right) + \mathrm{var}\left(z_{.2}c_{p2}\right) + 2\,\mathrm{cov}\left(z_{.1}c_{p1}, z_{.2}c_{p2}\right)$$

$$= c_{p1}^2\,\mathrm{var}\left(z_{.1}\right) + c_{p2}^2\,\mathrm{var}\left(z_{.2}\right) + 2\,\mathrm{cov}\,c_{p1}c_{p2}\left(z_{.1}, z_{.2}\right)$$

$$= c_{p1}^2 + c_{p2}^2$$

$$= \text{Communality of variable } p$$

Thus,

$$\mathrm{var}\left(d_{.p}\right) = \text{Communality} + \text{Error Variance}$$

$$= \text{Communality} + \text{Uniqueness}$$

Factor Analysis
Rotational Indeterminacy under Orthogonal rotations

$$\mathbf{R} = \mathbf{CC'} + <$$

$$\mathbf{R} = \mathbf{CT(CT)'} + \triangleleft$$

$$\mathbf{R} = \mathbf{CTT'C'} + \triangleleft$$

$$\mathbf{R} = \mathbf{CIC'} + \triangleleft$$

$$\mathbf{R} = \mathbf{CC'} + \triangleleft$$

Rotation via Varimax
Rotating the Principal Components Solution

Original PCA Loadings

	PC1	PC2
Murder	0.54	-0.42
Assault	0.58	-0.19
UrbanPop	0.28	0.87
Rape	0.54	0.17

2 PCA Loadings rotated via Varimax rotation

	PC1	PC2
Murder	0.659	-0.167
Assault	0.609	
UrbanPop		0.911
Rape	0.430	0.372

Rotation via Varimax
Rotating the Principal Components Solution

```
x=prcomp(USArrests,scale=TRUE)
summary(x)
cor(as.vector(scale(USArrests)), as.vector(x$x[,1:2]  %*% t(x$rotation)[1:2,]))^2

y=varimax(x$rotation[,1:2])
cor(as.vector(scale(USArrests)), as.vector(x$x[,1:2] %*% y$rotmat %*% t(y$rotmat) %*%
t(x$rotation)[1:2,]))^2
```

Factor Analysis

Any Questions?