GEOG606 - HW03

Anika Cartas

Tuesday, March 24, 2015

### Assignment Description

In 1898, Bumpus did a series study of English sparrows. Attached is his measurement of characteristics of 49 female sparrows:

X1=total length X2=alar length, X3=length of beak and head, X4=length of humerus, X5=length of keel and sternum.

Please use principle component analysis to find out the linear combination of 5 variables that best discriminates different sparrows. First, use the PCA routine in R to do it.

Second, try to reproduce the result of PCA routine by following the key steps of PCA:

* generate covariance matrix
* solve the eigen value equation to calculate the eigen values
* derive eigen vectors
* calculate loading vectors

### 1. Read in the data

data <- read.table("hw03.csv",header=TRUE, col.names=c("Total","Alar","Beak.Head","Humerus","Keel.Sternum"))  
head(data)

## Total Alar Beak.Head Humerus Keel.Sternum  
## 1 156 245 31.6 18.5 20.5  
## 2 154 240 30.4 17.9 19.6  
## 3 153 240 31.0 18.4 20.6  
## 4 153 236 30.9 17.7 20.2  
## 5 155 243 31.5 18.6 20.3  
## 6 163 247 32.0 19.0 20.9

### 2. Use PCA routine in R

prcomp.results <- prcomp(data)  
  
#Rotation Matrix  
prcomp.results

## Standard deviations:  
## [1] 5.9435475 2.1499905 0.7943033 0.5592706 0.2784261  
##   
## Rotation:  
## PC1 PC2 PC3 PC4 PC5  
## Total 0.53650052 -0.82809990 -0.15649065 -0.04020969 -0.01765243  
## Alar 0.82901535 0.55051223 -0.05774395 -0.06902156 0.03964203  
## Beak.Head 0.09649615 -0.03356237 0.23751487 0.89762653 0.35695288  
## Humerus 0.07435219 0.01459529 0.20324541 0.30724056 -0.92658150  
## Keel.Sternum 0.10030441 -0.09923405 0.93512262 -0.30575979 0.11021920

#Loading Scores  
head(prcomp.results$x)

## PC1 PC2 PC3 PC4 PC5  
## [1,] 1.9664223 3.689713 -0.16801141 0.06169590 0.16647852  
## [2,] -3.5023362 2.714180 -0.81488580 -0.49908927 0.03198139  
## [3,] -3.8434585 3.430206 0.52085909 -0.07244318 -0.08926600  
## [4,] -7.2613378 1.260991 0.21176257 0.02111595 0.32098994  
## [5,] -0.2503842 3.441451 -0.08648433 0.24206207 -0.04555039  
## [6,] 7.4958530 -1.051783 -0.80825616 0.03255193 -0.15422636

### 3. Reproduce results manually

#3.1 Zero data by subtracting mean   
data.adjusted <- apply(data,2,function(x) { x - mean(x) })   
  
#3.2 Generate covariance matrix  
cov.matrix <- cov(data.adjusted)  
cov.matrix

## Total Alar Beak.Head Humerus Keel.Sternum  
## Total 13.353741 13.610969 1.9220663 1.3306122 2.1922194  
## Alar 13.610969 25.682823 2.7136054 2.1977041 2.6578231  
## Beak.Head 1.922066 2.713605 0.6316327 0.3422662 0.4146471  
## Humerus 1.330612 2.197704 0.3422662 0.3184184 0.3393707  
## Keel.Sternum 2.192219 2.657823 0.4146471 0.3393707 0.9828231

#3.3 Solve the eigen value equation to calculate the eigen values  
cov.eigen <- eigen(cov.matrix)  
cov.eigen$values

## [1] 35.32575695 4.62245934 0.63091779 0.31278363 0.07752107

#3.4 Derive eigen vectors  
cov.eigen$vectors

## [,1] [,2] [,3] [,4] [,5]  
## [1,] -0.53650052 0.82809990 0.15649065 0.04020969 -0.01765243  
## [2,] -0.82901535 -0.55051223 0.05774395 0.06902156 0.03964203  
## [3,] -0.09649615 0.03356237 -0.23751487 -0.89762653 0.35695288  
## [4,] -0.07435219 -0.01459529 -0.20324541 -0.30724056 -0.92658150  
## [5,] -0.10030441 0.09923405 -0.93512262 0.30575979 0.11021920

#3.5 Calculate loading scores  
manual.results <- t(cov.eigen$vectors) %\*% t(as.matrix(data.adjusted))  
manual.results <- t(manual.results)  
head(manual.results)

## [,1] [,2] [,3] [,4] [,5]  
## [1,] -1.9664223 -3.689713 0.16801141 -0.06169590 0.16647852  
## [2,] 3.5023362 -2.714180 0.81488580 0.49908927 0.03198139  
## [3,] 3.8434585 -3.430206 -0.52085909 0.07244318 -0.08926600  
## [4,] 7.2613378 -1.260991 -0.21176257 -0.02111595 0.32098994  
## [5,] 0.2503842 -3.441451 0.08648433 -0.24206207 -0.04555039  
## [6,] -7.4958530 1.051783 0.80825616 -0.03255193 -0.15422636

#3.6. Compare Results  
head(prcomp.results$x)

## PC1 PC2 PC3 PC4 PC5  
## [1,] 1.9664223 3.689713 -0.16801141 0.06169590 0.16647852  
## [2,] -3.5023362 2.714180 -0.81488580 -0.49908927 0.03198139  
## [3,] -3.8434585 3.430206 0.52085909 -0.07244318 -0.08926600  
## [4,] -7.2613378 1.260991 0.21176257 0.02111595 0.32098994  
## [5,] -0.2503842 3.441451 -0.08648433 0.24206207 -0.04555039  
## [6,] 7.4958530 -1.051783 -0.80825616 0.03255193 -0.15422636

#Calculate differences between each value and see if any are not nearly identical.  
table(abs(manual.results) - abs(prcomp.results$x) < 0.001)

##   
## TRUE   
## 245