assignment08

Jason Grahn

3/5/2019

# Problem 13.7(a)

## 13.7 Use the words data of Table 5.9.

### (a) Obtain principal component loadings for two factors.

essay <- read.table(here::here("assignment08/T5\_9\_ESSAY.DAT")) %>%   
 rename(student = V1,  
 y1 = V2, # informal\_words  
 y2 = V3, # informal\_verbs  
 x1 = V4, # formal\_words  
 x2 = V5) %>% # formal\_verbs  
 select(y1:x2)  
  
head(essay,3)

## y1 y2 x1 x2  
## 1 148 20 137 15  
## 2 159 24 164 25  
## 3 144 19 224 27

# obtain the corralation matrix of the essay data  
corMax <- cor(essay)  
corMax

## y1 y2 x1 x2  
## y1 1.0000000 0.7660725 0.5953551 0.2173378  
## y2 0.7660725 1.0000000 0.5600505 0.4427548  
## x1 0.5953551 0.5600505 1.0000000 0.7202028  
## x2 0.2173378 0.4427548 0.7202028 1.0000000

#obtain the eigenvalues from the correlation matrix  
eigenValue <- eigen(corMax)  
eigenValue

## eigen() decomposition  
## $values  
## [1] 2.6657459 0.8993358 0.3276382 0.1072801  
##   
## $vectors  
## [,1] [,2] [,3] [,4]  
## [1,] -0.4914201 0.5642628 -0.3208298 0.5806737  
## [2,] -0.5241023 0.3441774 0.6604771 -0.4130722  
## [3,] -0.5409202 -0.2848265 -0.6020434 -0.5136371  
## [4,] -0.4372967 -0.6942790 0.3136590 0.4778769

# keep only the first two columns of vectors into matrix "C"  
## we're keeping only the first two columns because that's the intent of the question  
C <- eigenValue$vectors[,1:2]   
C

## [,1] [,2]  
## [1,] -0.4914201 0.5642628  
## [2,] -0.5241023 0.3441774  
## [3,] -0.5409202 -0.2848265  
## [4,] -0.4372967 -0.6942790

# keep only the first two eigenvalues as matrix "D1"  
D1 <- diag(eigenValue$values[1:2],   
 nrow =2,   
 ncol =2)  
D1

## [,1] [,2]  
## [1,] 2.665746 0.0000000  
## [2,] 0.000000 0.8993358

#build loadings from C and D1  
loadings <- C %\*% (D1^(1/2))  
loadings

## [,1] [,2]  
## [1,] -0.8023471 0.5351091  
## [2,] -0.8557077 0.3263949  
## [3,] -0.8831664 -0.2701104  
## [4,] -0.7139792 -0.6584078

Factor loadings are correlations between the factors and each of the variables. Reviewing the factor loadings for the dataset shows that Factor 1 has a strong relationships with ALL variables and can be considered practically significant across the board. When we review factor 2, we see some practical significance, however, we also see much weaker correlation for the and variables informal\_verbs and formal\_words.