STAT-S675

Homework 5

John Koo

Problem 1

[Exercise 4.6.1 in the notes]

Want to show $\tau(\kappa(\Gamma)) = P\Gamma P$.

$P = (I - \frac{ee^T}{n})$
$\tau(X) = -\frac{1}{2}PXP$
$\kappa(X) = \text{diag}(X)e^T - 2X + e\text{diag}(X)^T$

Since Γ is a similarity matrix, $\text{diag}(\Gamma) = e$ (we can just normalize everything to this WLOG).

$$\begin{aligned}\kappa(\Gamma) &= \text{diag}(\Gamma)e^T - 2\Gamma + e\text{diag}(\Gamma)^T \\ &= ee^T - 2\Gamma + ee^T \\ &= 2(ee^T - \Gamma)\end{aligned}$$

So if we want to find $\tau(\kappa(\Gamma))$:

$$\begin{aligned}\tau(\kappa(\Gamma)) &= \tau(2(ee^T - \Gamma)) \\ &= -\frac{1}{2}P(2(ee^T - \Gamma))P \\ &= P(\Gamma - ee^T)P \\ &= P\Gamma P - Pee^T P\end{aligned}$$

So we just need to show that $Pee^T P = 0$. First, note that $e^T e = n$.

$$\begin{aligned}Pee^T P &= (I - \frac{ee^T}{n})ee^T(I - \frac{ee^T}{n}) \\ &= (Iee^T - \frac{ee^T ee^T}{n})(I - \frac{ee^T}{n}) \\ &= (ee^T - \frac{nee^t}{n})(I - \frac{ee^T}{n}) \\ &= (ee^T - ee^T)(I - \frac{ee^T}{n}) \\ &= 0\end{aligned}$$

Therefore $\tau(\kappa(\Gamma)) = P\Gamma P$.

Problem 2

[Exercise 4.6.3 in the notes]

```r
import::from(readr, read_table2)
import::from(magrittr, `%>%`, `%<>%`)
library(ggplot2)
import::from(GGally, ggpairs)
import::from(ggrepel, geom_label_repel)

# read the data
Delta <- read_table2('http://pages.iu.edu/~mtrosset/Courses/675/congress.dat',
                     col_names = FALSE) %>%
  as.matrix()

# code from the notes
Delta.cmds <- cmdscale(Delta, k = 3, eig = TRUE)
X <- Delta.cmds$points
```

Part a

```r
r <- sum(Delta.cmds$eig > 0)
sum.pos.eig.vals <- sum(sapply(Delta.cmds$eig, function(i) max(i, 0)))
sum.neg.eig.vals <- sum(sapply(Delta.cmds$eig, function(i) min(i, 0)))
```

9 out of 15 eigenvalues are positive. Using the

Part b

```r
sum(sort(Delta.cmds$eig, decreasing = TRUE)[1:2]) / sum.pos.eig.vals
```

```
[1] 0.6998389
```

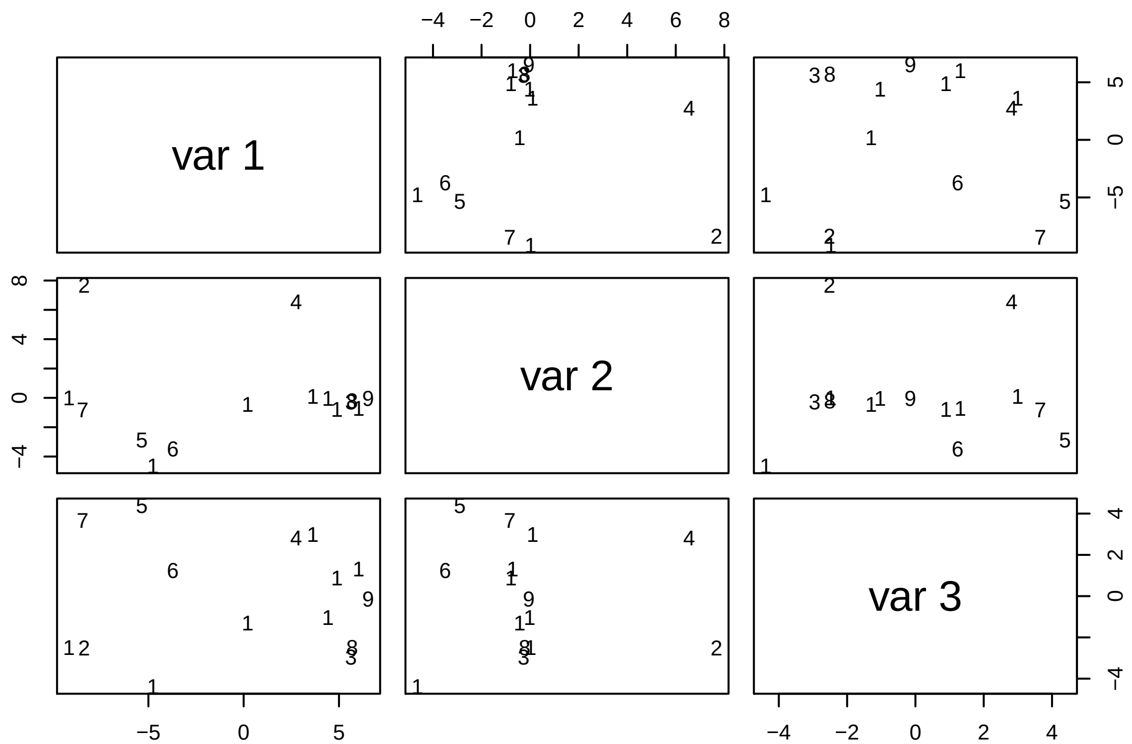
```r
sum(sort(Delta.cmds$eig, decreasing = TRUE)[1:3]) / sum.pos.eig.vals
```

```
[1] 0.8116862
```

Two eigenvalues is ~70% of the sum of the positive eigenvalues while three is ~80%.

Part c

```r
pairs(X, pch = as.character(seq(15)))
```




```r
X.df <- X %>%
  as.data.frame() %>%
  dplyr::mutate(id = rownames(.))
```