Practice Exam 05 PCA and MDS Visualisation

Ryan Greenup

June 9, 2020

Contents

Trends 2014 Q8

Take the Following Data:

	Period 1	Period 2	Period 3
Day 1	66	97	69
Day 2	68	111	87
Day 3	103	112	87

In order to examine the fluctuation for each day, break the data into its trend and periodic components, after a square root transformation:

```
data <- rbind(
  c(66,  97,  69),
  c(68,  111, 87),
  c(103, 112, 87)
)
sqrt(data)
```

```
         [,1]      [,2]     [,3]
[1,]  8.124038  9.848858 8.306624
[2,]  8.246211 10.535654 9.327379
[3,] 10.148892 10.583005 9.327379
```

In this case the period is a time of day, so for example Morning, Lunch, Evening, so a seasonal trend would be expected.

Taking a moving average over the period of that trend however will remove the influence of that seasonality:

$$\begin{aligned} Y = T + S + & \\ &= T + 0 + 0 \\ &= T \end{aligned}$$

(a) Calculate the Trend

Remember that the moving average the window expands evenly in both directions, so for example day2, period2 would be calculated as:

```
library(tidyverse)
c(68, 11, 87)  %>%
  sqrt()  %>%
  sum() / 3
```

```
[1] 6.963405
```

(b) Calculate the Periodic Component

The periodic component is S and can be calculated:

$$\begin{aligned} Y = T + S + & \\ S = Y - T - & \end{aligned}$$

but this doesn't sum to zero so instead:

$$S = S' - 1/mS'$$

So the seasonal trend would be:

```
library(zoo)
Y <- data

T <- rollmean(sqrt(as.vector(data)), k = 3)
Y_trim <- sqrt(head(as.vector(data), -1)[-1])
Sp <- Y_trim - T
S <- Sp - mean(Sp)

matrix(c("X", S, "X"), nrow = 3)
```

```
     [,1]                 [,2]                  [,3]
[1,] "X"                  "-0.334760992717553" "-1.10486327843698"
[2,] "-0.599320243685816" "0.207330398617034"  "0.334433975456945"
[3,] "0.728420271117701"  "0.768759869648668"  "X"
```

Trend

	Period 1	Period 2	Period 3
Day 1	NA	8.76	8.8
Day 2	9.03	6.96	10
Day 3	10.02	10.02	NA

Periodic

Subtact the Trend from the observations:

S = Y - T

	Period 1	Period 2	Period 3
Day 1	NA	8.76-sqrt(97)	8.8-sqrt(69)
Day 2	sqrt(68) - 9.03	sqrt(111) - 6.96	sqrt(87) - 10
Day 3	sqrt(103) - 10.02	sqrt(112) - 10.02	NA

Now take the column averages

```
sp <-
c(

mean(c(sqrt(68) - 9.03, sqrt(103) - 10.02)),

mean(c(sqrt(97) - 8.76,
 sqrt(111) - 6.96,
 sqrt(112) - 10.02)),


mean(c(sqrt(69) - 8.8,
  sqrt(87) - 10 ))

)
sp
```

```
[1] -0.3274486  1.7425056 -0.5829985
```

Unfourtunately this does not sum to zero so we fix that:

```
(S <- sp - mean(sp))
```

```
[1] -0.6048014  1.4651528 -0.8603514
```

This does not match the sheet but I don't understand why?

In practice

In practice we would just conclude that the values must sum to zero:

```
p1 <- - 0.336
p3 <- - 0.594
# 0 = p1 + p2 + p3
(p2 = -p1 - p3)
```

```
[1] 0.93
```

Thus the missing componente is 0.93