

Assignment 2

Instructor: Jiguo Cao

Instructions: It is due at 5pm, Monday, March 18, 2019.

1. [Rangel et al.(2004)Rangel, Angus, Ghahramani, Lioumi, Sotheran, Gaiba, Wild, and Falciani] collected time course microarray data for T-cell activation to investigate a dynamic gene regulatory network. For the time course measurement of each gene, 10 unequally spaced observations are sampled; each gene has 34 independent replicates. Please select one gene and analyze the 34 replicates of the time course measurements of the selected gene. Figure 1 displays the gene expression profiles at each sampled point across the 34 replicates for Gene F10. The R codes to load the data are provided below.

```
install.packages("fda")
library(fda)

#provide working directory of the data file
Gene <- read.table("/Users/cao/Dropbox/Teaching/FDA/SummerCourse2018/Assignments/
  Assignment2/gene58rep34.txt",header=T)
head(Gene)
unique(Gene$gene)
table(Gene$gene)

#=====
#take F10 out
gene.10 <- Gene[Gene$gene=="F10",]
head(gene.10)
dim(gene.10)

# observation time points
timepts=seq(1,10,1)
# open a new graph window in R in a Windows laptop.
# run quartz() if you are using a Mac computer
windows()
matplot(t(gene.10[,2:11]),type="l",xlab="Time point", ylab="Gene expression")
```

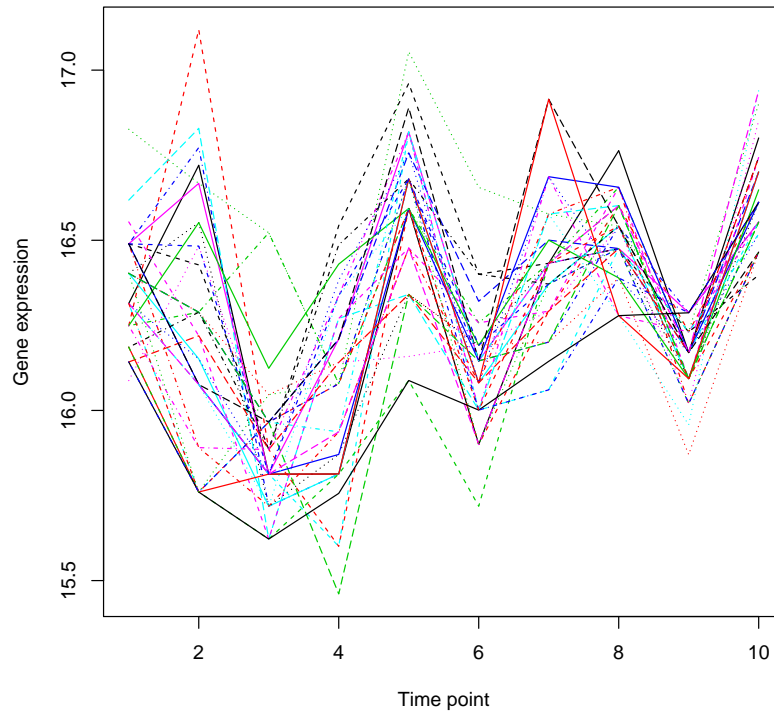


Figure 1: The gene expression of one selected gene.

- (a) (10') Can you use functional principal component analysis to explore the major modes of variability among these 34 curves?
- (b) (10') Can you clustering the 34 curves based on their functional principal component scores?

Please organize the above estimation results, and your R codes in a single pdf report.

References

- [Rangel et al.(2004)Rangel, Angus, Ghahramani, Lioumi, Sotheran, Gaiba, Wild, and Falciani]
 Rangel, C., Angus, J., Ghahramani, Z., Lioumi, M., Sotheran, E., Gaiba, A., Wild, D. L., and Falciani, F. (2004). Modeling T-cell activation using gene expression profiling and state-space models. Bioinformatics, **20**(9), 1361–1372.