

Problem Set 4

Shiyin Wang

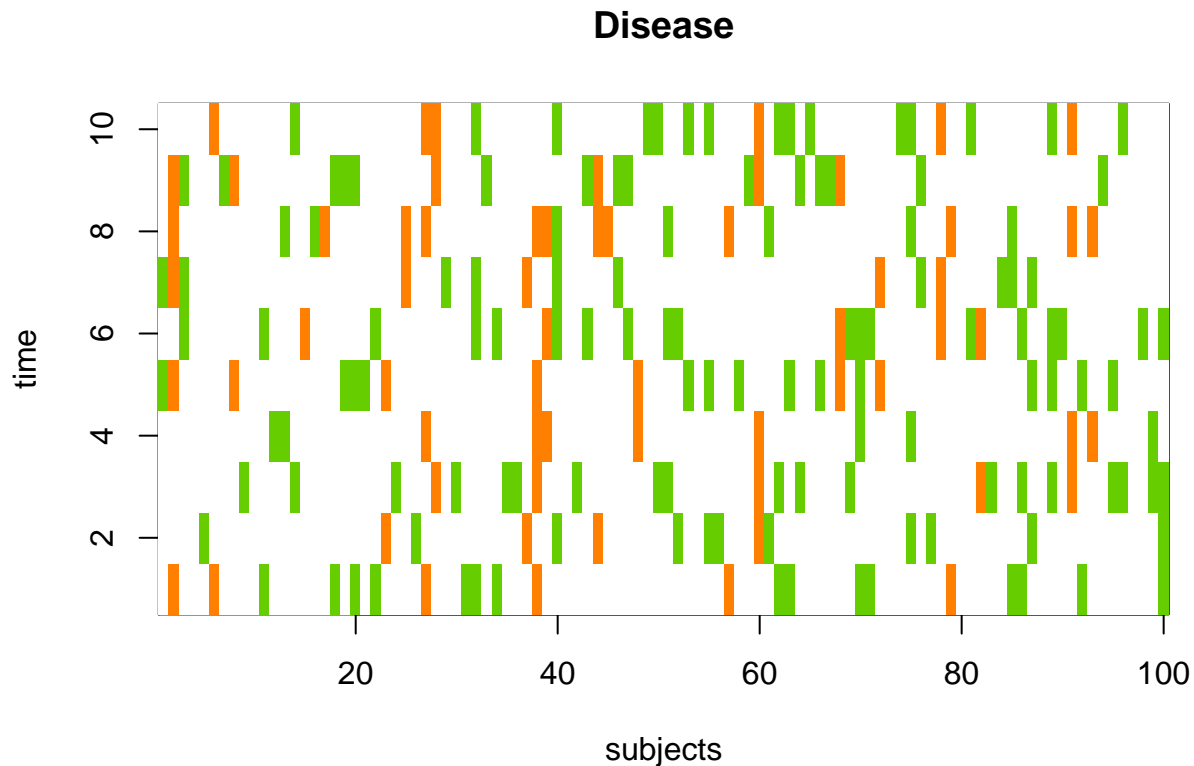
Problem 4.2 (g)

```
data <- read.table("~/Desktop/3_2/6.437/PS/ps4-data/data.txt", quote="", comment="")
mu = data[, 0:10]
y = data[, 11]
mu = as.matrix(mu)
y = as.list(y)
```

Data Visualization

I color all the exposures of infected subjects(columns) with red. The exposures of uninfected subjects are colored with green.

```
visual = mu
for(i in seq(1, 100, 1)){
  for(j in seq(1, 10, 1)){
    if(y[i]==1 && mu[i, j]==1)
      visual[i, j] = 2
  }
}
cols <- c("#FFFFFF", "chartreuse3", "#FF8000FF")
image(1:nrow(mu), 1:ncol(mu), visual, col=cols, xlab="subjects", ylab="time", main="Disease")
```



EM Algorithm

```
T = 5000
z = matrix(0.1, nrow=100, ncol=10)
r = rep(0.5, 10)

for(t in seq(1,T, 1)){
  mur = mu%*%diag(r)
  mur1 = 1 - mur
  prod = matrix(1, 100, 1)
  for(i in seq(1, 10, 1)){
    prod = prod * mur1[,i]
  }
  z1 = 1 - prod #Denominator of z
  z = diag(as.list(1/z1)) %*% mur
  for(i in seq(1, 100, 1)){
    if(y[i]==0){
      z[i, ] = matrix(0, 1, 10)
    }
  }
  r = colSums(mu * z)/colSums(mu)
}

print(r)
```

| ## | V1 | V2 | V3 | V4 | V5 |
|----|---------------|--------------|---------------|--------------|--------------|
| ## | 4.940656e-324 | 1.915936e-02 | 9.881313e-324 | 2.619160e-01 | 2.057405e-01 |
| ## | V6 | V7 | V8 | V9 | V10 |
| ## | 1.208490e-01 | 1.468435e-01 | 5.566214e-01 | 5.622997e-64 | 1.547256e-01 |