Final Project

Xinyi Lin

5/10/2019

```
library(tidyverse)
library(lubridate)
library(truncnorm)
```

Problem 1

Question 1

First, we need to tidy data. As the day of year $x_{i,1}(t)$ and time t are two variables in given model, we need to transform `time` into some suitable formats. Normally, records are taken every 6 hour at 00:00:00, 06:00:00, 12:00:00 and 18:00:00. However, some records are taken at other timepoints. These records are ineffective as there are no records 6 hour before or after those timepoints to help train or test given models. Thus, we remove those observations.

```
shift <- function(x, n=1){
  c(x[-(seq(n))], rep(NA, n))
}
```

```
data1 = read.csv("./hurrican356.csv") %>%
  janitor::clean_names() %>%
  mutate(year = season,
         date_hour = time) %>%
  separate(date_hour, into = c("date", "hour"), sep = " ") %>%
  filter(hour == "00:00:00)" | hour == "06:00:00)" | hour == "12:00:00)" | hour == "18:00:00)") %>%
  mutate(hour = str_replace(hour, ":00:00\\)", ""),
         hour = as.numeric(hour),
         date = str_replace(date, "\\(", ""),
         date = yday(date),
         nature = as.numeric(as.factor(nature))) %>%
  group_by(id) %>%
  mutate(delta1 = c(NA, diff(latitude)),
         delta2 = c(NA, diff(longitude)),
         delta3 = c(NA, diff(wind_kt)),
         latitude_p = shift(latitude),
         longitude_p = shift(longitude),
         windkt_p = shift(wind_kt)) %>%
  ungroup() %>%
  na.omit() %>%
  select(id, latitude, longitude, wind_kt, latitude_p, longitude_p, windkt_p, date, year, nature, delta

#head(data1)
#summary(data1)
```

Then, we randomely select 80% hurricanes.

```
set.seed(123)
id = unique(data1$id)
```

```
num_id = length(id)
train_id = sample(id, 0.8*num_id)

train_data = data1[which(data1$id %in% train_id),] %>%
  select(-id)
```

After that, we use M-H algorithm to develop an MCMC process.

As

$$Y_{i,j}(t+6) = \mu_{i,j}(t) + \rho_j Y_{i,j}(t) + \epsilon_{i,j}(t)$$

$$Y_i \sim MVN\left(\begin{bmatrix} \mu_{i,1}(t) + \rho Y_{i,1}(t) \\ \mu_{i,2}(t) + \rho Y_{i,2}(t) \\ \mu_{i,3}(t) + \rho Y_{i,3}(t) \end{bmatrix}, \Sigma \right)$$

$$f_{Y_i(t+6)}(Y_i|\rho_j, \beta, \Sigma) = \frac{exp[-\frac{1}{2}(Y_i-\mu)^T \Sigma^{-1}(Y_i-\mu)]}{\sqrt{(2\pi)^k|\Sigma|}}$$

$$\pi(\theta) = \prod \frac{exp[-\frac{1}{2}(Y_i-\mu)^T \Sigma^{-1}(Y_i-\mu)]}{\sqrt{(2\pi)^k|\Sigma|}} \times \pi_1(\beta)\pi_2(\rho_1)\pi_3(\rho_2)\pi_4(\rho_3)\pi_5(\Sigma^{-1})$$

$$\pi'(\theta) = \sum log[\frac{exp[-\frac{1}{2}(Y_i-\mu)^T \Sigma^{-1}(Y_i-\mu)]}{\sqrt{(2\pi)^k|\Sigma|}}] + log[\pi_1(\beta)] + log[\pi_2(\rho_1)] + log[\pi_3(\rho_2)] + log[\pi_4(\rho_3)] + log[\pi_5(\Sigma^{-1})]$$

Let $\theta = (\beta, \rho_j, \Sigma)$, with probability

$$\alpha(\lambda|\theta^{(t)}) = min\left\{1, \frac{\pi'(\lambda)q(\theta^{(t)}|\lambda)}{\pi'(\theta^{(t)})q(\lambda|\theta^{(t)})}\right\}$$

Accept $\theta^{(t+1)} = \lambda$, else, set$\theta^{(t+1)} = \theta^{(t)}$.

Starting points.

```
set.seed(123)
# rho: 1*3 vector
#rho = rtruncnorm(3, a=0, b=1, mean = 0.5, sd = 1/5)
rho = rep(0.9, 3)
# epsilon: 1*3 vector
sigma = bayesm::rwishart(3,diag(0.1,3))$IW
sigma = c(sigma[1,], sigma[2,c(2,3)], sigma[3,3])
#rWishart(1,3,diag(0.1,3))
#means = c(0,0,0)
#epsilon = mvtnorm::rmvnorm(1, means, sigma)
#beta: 1*18 vector, where beta_kj is the [6*(j-1)+(k+1)]th number.
#beta = mvtnorm::rmvnorm(1, rep(0,21), diag(1,21))
beta = rep(0.005,21)
```

Density function.

```
# for each yi
logdy = function(obs, beta.=beta, rho.=rho, sigma.=sigma){
  x = c(1,obs[7:12])
  y = obs[1:3]
```

```
  beta_m = matrix(beta.,3)
  sigma_m = matrix(c(sigma.[c(1:3)], sigma.[2], sigma.[c(4,5)], sigma.[c(3,5)], sigma.[6]), 3)
  mu = beta_m %*% x + rho.*obs[1:3]
  dy = mvtnorm::dmvnorm(obs[4:6], mean = mu, sigma = sigma_m)
  return(log(dy))
}

#traintest = train_data[c(1:1000),]
#apply(traintest, 1, mu)
```