Xinru Wang

xw2676

2020/5/6

```r
data=read.csv("hurrican356.csv") %>% janitor::clean_names()

## select 80% hurricanes as train data set
hurricanes = unique(as.character(data$id))
trRows <- createDataPartition(1:356,
                              p = 0.8,
                              list = FALSE)
train_hurricanes = hurricanes[trRows]
test_hurricanes = hurricanes[-trRows]

df=data %>% separate(time,into=c("sign","date","hour","sign2"),sep = c(1,9,18)) %>% dplyr::select(-sign
  filter(hour == " 00:00:00" | hour == " 06:00:00" | hour == " 12:00:00" | hour == " 18:00:00") %>%
  group_by(id)  %>%
  mutate(
    date=as.Date(date, '%y-%m-%d'),
    begin=str_c(as.character(season),'0101'),
    initial=as.Date(begin, "%Y%m%d"),
    days=as.numeric(date-initial),
    final=str_c(date,hour),
    time=as.numeric(difftime(final,final[1],units = "hour")),
    nature = as.numeric(nature)-1
  ) %>% dplyr::select(-begin,-initial)

## DS ET NR SS TS

df_train = df %>%
  filter(id %in% train_hurricanes) %>%
  mutate(
    shift_lat = shift(latitude, fill=NA,type="lag"),
    diff_lat = latitude - shift_lat,
    shift_long = shift(longitude, fill=NA,type="lag"),
    diff_long = longitude - shift_long,
    shift_wind = shift(wind_kt, fill=NA,type="lag"),
    diff_wind = wind_kt - shift_wind,
  ) %>%
  drop_na()


df_test = df %>%
  filter(id %in% test_hurricanes)  %>%
  mutate(
    shift_lat = shift(latitude, fill=NA,type="lag"),
```

```
    diff_lat = latitude - shift_lat,
    shift_long = shift(longitude, fill=NA,type="lag"),
    diff_long = longitude - shift_long,
    shift_wind = shift(wind_kt, fill=NA,type="lag"),
    diff_wind = wind_kt - shift_wind,
  ) %>%
  drop_na()

## time, days and id are created useful variables
```
```
loglikelihood = NULL
loglikeli_func = function(dat,beta,sigma, rho){
  dat = dat %>%
    mutate(
      shift_wind_2 = shift(wind_kt),
      mu = beta[1] + days*beta[2] + season*beta[3] + nature*beta[4] + diff_lat*beta[5] + diff_long*beta
      loglikeli = log(dnorm(wind_kt, mean = mu, sd = sigma))
    ) %>%
    drop_na()

  loglikelihood = sum(dat$loglikeli) + log(dmvnorm(beta, rep(0,7), diag(1,7))) + log(truncnorm(rho, a=0
  return(loglikelihood)
}
```
```
prop <- 1 + (runif(1) - 0.5) * 2
prop
```
```
## [1] 0.8748402
```