Second Order

Ricky Trujillo

2024-05-09

```
weather_summary = read.csv("weather.csv")
weather_summary$weather_state = as.factor(weather_summary$weather_state)

states <- c("Chilly", "Cloudy", "Partly Cloudy", "Rainy", "Sunny")
weather_summary$prediction<- rep(NA,nrow(weather_summary))
weather_summary$prediction = factor(weather_summary$prediction, levels= states)


weather24_pred = weather_summary %>%
  filter(Date>= as.Date("2024-01-01", "%Y-%m-%d"))%>%
  select(Date, Month, weather_state, prediction)

state_mat = one_hot(as.data.table(weather_summary$weather_state))
colnames(state_mat)<-  states
```

```
# Matrix A: One-Step Transition Matrix of k-1 state
A = matrix(NA, nrow = 5, ncol = 5, dimnames = list(states,states))

for (i in states){ # c("Chilly", "Cloudy", "Partly Cloudy", "Rainy", "Sunny")
  for(j in states){ # c("Chilly", "Cloudy", "Partly Cloudy", "Rainy", "Sunny")
    count_1a=0
    count_2a=0
    for(k in 2:nrow(weather_summary)){
      if(weather_summary$weather_state[k-1][1]==j){
        count_2a = count_2a +1
      if(weather_summary$weather_state[k][1]==i){
        count_1a = count_1a +1
      }
    }
    }
    A[i,j] = count_1a/count_2a
  }
}

round(A,2)
```

```
##               Chilly Cloudy Partly Cloudy Rainy Sunny
## Chilly          0.68   0.26          0.02  0.46  0.08
## Cloudy          0.18   0.49          0.34  0.24  0.10
## Partly Cloudy   0.01   0.06          0.20  0.05  0.09
## Rainy           0.03   0.08          0.00  0.24  0.01
## Sunny           0.10   0.11          0.44  0.00  0.73
```

```r
colSums(A)
```

```
##        Chilly        Cloudy Partly Cloudy         Rainy         Sunny
##             1             1             1             1             1
```

```r
# Matrix B: One-Step Transition Matrix of k-2 state
B = matrix(NA, nrow = 5, ncol = 5, dimnames = list(states,states))

for (i in states){
  for(j in states){
    count_1b=0
    count_2b=0
    for(k in 3:nrow(weather_summary)){
      if(weather_summary$weather_state[k-2][1]==j){
        count_2b = count_2b +1
      if(weather_summary$weather_state[k][1]==i){
        count_1b = count_1b +1
      }
    }
    }
    B[i,j] = count_1b/count_2b
  }
}

round(B,2)
```

```
##               Chilly Cloudy Partly Cloudy Rainy Sunny
## Chilly          0.57   0.34          0.08  0.51  0.12
## Cloudy          0.25   0.36          0.30  0.24  0.14
## Partly Cloudy   0.01   0.05          0.16  0.03  0.10
## Rainy           0.03   0.09          0.00  0.19  0.01
## Sunny           0.14   0.16          0.46  0.03  0.63
```

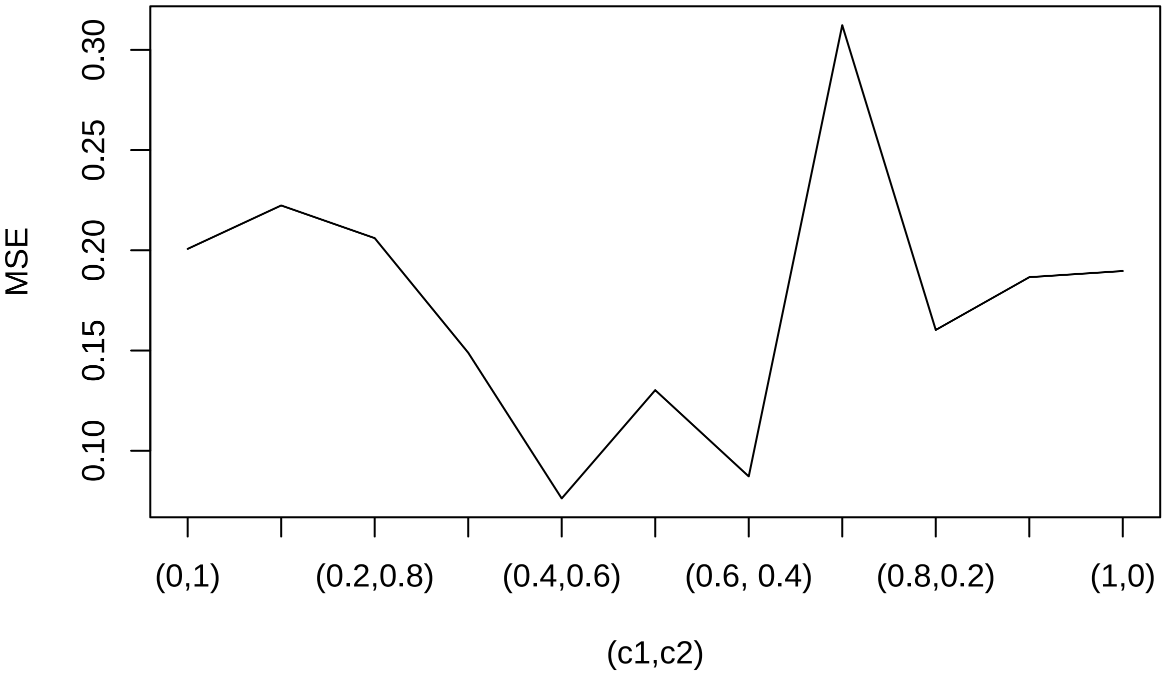
```r
colSums(B)
```

```
##        Chilly        Cloudy Partly Cloudy         Rainy         Sunny
##             1             1             1             1             1
```

```r
second_order <- function(weather24_pred, c1,c2){
  prob = list()
  for(i in 1:nrow(weather24_pred)){
    if(i==1){
      prob[[i]] = c1*A %*% t(tail(state_mat, 1)) + c2*B %*% t(slice(state_mat, n()-2))
      choice = sample(1:5,1,prob=prob[[i]])
      weather24_pred$prediction[i] = states[choice]
    }else if(i==2){
      state_bin1 = one_hot(as.data.table(factor(weather24_pred$prediction[i-1], levels = states)))
      prob[[i]] = c1*A %*% t(state_bin1) + c2*B %*% t(slice(state_mat, n()-1))
      choice = sample(1:5,1,prob=prob[[i]])
      weather24_pred$prediction[i] = states[choice]
    }else{
      state_bin1 = one_hot(as.data.table(factor(weather24_pred$prediction[i-1], levels = states)))
      state_bin2 = one_hot(as.data.table(factor(weather24_pred$prediction[i-2], levels = states)))
      prob[[i]] = c1*A %*% t(state_bin1) + c2*B %*% t(state_bin2)
      choice = sample(1:5,1,prob=prob[[i]])
      weather24_pred$prediction[i] = states[choice]
    }
  }
  return(weather24_pred)
}
```

```
set.seed(1111)
mse<-function(x_hat,x) rowMeans((x_hat-x)^2)
someData <- rep(NA, 4*5)  #(4 Months, 5 States, S simulations)
avg_state <- array(someData, c(4, 5))

# store MSE for each pair (c1,c2)
MSE = list()
# make the original matrix of ratios for each month (manually)
weather_ratio = matrix(c(0.58064516, 0.25806452, 0.06451613, 0.06451613,0.03225806,
                         0.41379310, 0.34482759, 0, 0.24137931,0,
                         0.54838710, 0.29032258, 0, 0.12903226, 0.03225806,
                         0.6, 0.33333333, 0, 0.03333333, 0.03333333), nrow = 4, ncol = 5, byrow = TRUE)

  m=1
  c1 = 0
  c2 = 1-c1
  while(c1<=1){
    weather24_pred$prediction<- rep(NA,nrow(weather24_pred))
    temp = second_order(weather24_pred, c1, c2)
    temp_avg = temp %>% group_by(Month) %>% mutate(month_len = length(Month)) %>%
    ungroup() %>%
    group_by(Month, prediction) %>%
    reframe(ratio = length(prediction)/month_len)%>%
    unique()
  for(i in 1:nrow(temp_avg)){
    if(temp_avg$Month[i]==1){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[1,1] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[1,2] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[1,3] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[1,4] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[1,5] = temp_avg$ratio[i]
      }
    }else if(temp_avg$Month[i]==2){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[2,1] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[2,2] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[2,3] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[2,4] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[2,5] = temp_avg$ratio[i]
      }
```

```r
    }else if(temp_avg$Month[i]==3){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[3,1] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[3,2] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[3,3] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[3,4] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[3,5] = temp_avg$ratio[i]
      }
    }else if(temp_avg$Month[i]==4){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[4,1] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[4,2] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[4,3] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[4,4] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[4,5] = temp_avg$ratio[i]
      }
    }
  }
  avg_state[is.na(avg_state)] <- 0

  mse_cpair= mse(avg_state,weather_ratio)
  mse_total = sum(mse_cpair)
  mse_total
  # Store MSE to list
  MSE[[m]] = mse_total
  c1=c1+0.1
  m=m+1
  }

metric_mat = matrix(unlist(MSE), nrow = 1, ncol = 11)
colnames(metric_mat) = c("(0,1)", "(0.1,0.9)", "(0.2,0.8)",
                     "(0.3,0.7)", "(0.4,0.6)", "(0.5,0.5)",
                     "(0.6, 0.4)", "(0.7,0.3)", "(0.8,0.2)",
                     "(0.9,0.1)", "(1,0)")

plot(metric_mat[1, ], type="l", xaxt='n',
     xlab="(c1,c2)", ylab="MSE")
axis(side=1,at=1:11,labels= c("(0,1)", "(0.1,0.9)", "(0.2,0.8)",
                     "(0.3,0.7)", "(0.4,0.6)", "(0.5,0.5)",
                     "(0.6, 0.4)", "(0.7,0.3)", "(0.8,0.2)",
                     "(0.9,0.1)", "(1,0)"))
```



```r
#Steady State Distribution
set.seed(1111)
prob = list()
c1 = 0.4
c2 = 0.6

  for(i in 1:1000){
    if(i==1){
      prob[[i]] = c1*A %*% t(tail(state_mat, 1)) + c2*B %*% t(slice(state_mat, n()-2))
    }else if(i==2){
      prob[[i]] = c1*A %*% prob[[i-1]] + c2*B %*% t(slice(state_mat, n()-1))
    }else{
      prob[[i]] = c1*A %*% prob[[i-1]] + c2*B %*% prob[[i-2]]
    }
  }
print(prob[[1000]])
```

```
##                     [,1]
## Chilly        0.34300638
## Cloudy        0.24372146
## Partly Cloudy 0.05887170
## Rainy         0.04356141
## Sunny         0.31083905
```

```r
S = 50# Number of Simulations
someData <- rep(NA, 4*5*S)  #(Months, States, S)
avg_state <- array(someData, c(4, 5, S))

for(s in 1:S){
  weather24_pred$prediction<- rep(NA,nrow(weather24_pred))
  temp = second_order(weather24_pred, c1=0.4, c2=0.6)
  temp_avg = temp %>% group_by(Month) %>% mutate(month_len = length(Month)) %>%
    ungroup() %>%
    group_by(Month, prediction) %>%
    reframe(ratio = length(prediction)/month_len)%>%
    unique()
  for(i in 1:nrow(temp_avg)){
    if(temp_avg$Month[i]==1){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[1,1,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[1,2,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[1,3,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[1,4,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[1,5,s] = temp_avg$ratio[i]
      }
    }else if(temp_avg$Month[i]==2){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[2,1,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[2,2,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[2,3,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[2,4,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[2,5,s] = temp_avg$ratio[i]
      }
    }else if(temp_avg$Month[i]==3){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[3,1,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[3,2,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[3,3,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[3,4,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[3,5,s] = temp_avg$ratio[i]
      }
    }else if(temp_avg$Month[i]==4){
      if(temp_avg$prediction[i]=="Chilly"){
        avg_state[4,1,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Cloudy"){
        avg_state[4,2,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Partly Cloudy"){
        avg_state[4,3,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Rainy"){
        avg_state[4,4,s] = temp_avg$ratio[i]
      }else if(temp_avg$prediction[i]=="Sunny"){
        avg_state[4,5,s] = temp_avg$ratio[i]
      }
    }
  }
  avg_state[is.na(avg_state)] <- 0
}
```

```
someData <- rep(0, 4*5)
sum_mat = array(someData, c(4, 5))
for(s in 1:S){
  sum_mat = sum_mat + avg_state[,,s]
}

sum_mat = sum_mat/S
round(sum_mat,2)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 0.37 0.26 0.05 0.06 0.25
## [2,] 0.33 0.24 0.07 0.04 0.32
## [3,] 0.34 0.23 0.07 0.04 0.33
## [4,] 0.34 0.24 0.08 0.04 0.31
```

```
sum(mse(sum_mat, weather_ratio))
```

```
## [1] 0.1124476
```