Assignment Probability & Stochastic Process

Getrude Gichuhi

2022-07-21

install.packages("ChannelAttribution") install.packages("ggplot2") install.packages("reshape") install.packages("dplyr") install.packages("plyr") install.packages("reshape2") install.packages("markovchain") install.packages("plotly") install.packages("purrrlyr")

Question 1 In the Dark Ages, Harvard, Dartmouth, and Yale admitted only male students. Assume that, at that time, 60 percent of the sons of Harvard men went to Harvard and the rest went to Yale, 30 percent of the sons of Yale men went to Yale, and the rest split evenly between Harvard and Dartmouth; and of the sons of Dartmouth men, 50 percent went to Dartmouth, 30 percent to Harvard, and 20 percent to Yale.

```
#Importing Libraries
library("ChannelAttribution")
```

```
## Warning: package 'ChannelAttribution' was built under R version 4.1.3

## ChannelAttribution 2.0.5

## Looking for attribution at path level? Try ChannelAttributionPro! Visit https://channelattribution.i
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
library("reshape")
```

```
## Warning: package 'reshape' was built under R version 4.1.3
```

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.1.3

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:reshape':
##
##     rename
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library("rlang")
```

```
## Warning: package 'rlang' was built under R version 4.1.3
```

```r
library("igraph")
```

```
## Warning: package 'igraph' was built under R version 4.1.3

##
## Attaching package: 'igraph'

## The following object is masked from 'package:rlang':
##
##     is_named

## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```r
library("plyr")
```

```
## Warning: package 'plyr' was built under R version 4.1.3

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:reshape':
##
##     rename, round_any
```

```
library("reshape2")
```

```
## Warning: package 'reshape2' was built under R version 4.1.3

##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:reshape':
##
##     colsplit, melt, recast
```

```
library("markovchain")
```

```
## Warning: package 'markovchain' was built under R version 4.1.3

## Package:  markovchain
## Version:  0.9.0
## Date:     2022-07-01
## BugReport: https://github.com/spedygiorgio/markovchain/issues
```

```
library("plotly")
```

```
## Warning: package 'plotly' was built under R version 4.1.3

##
## Attaching package: 'plotly'

## The following objects are masked from 'package:plyr':
##
##     arrange, mutate, rename, summarise

## The following object is masked from 'package:igraph':
##
##     groups

## The following object is masked from 'package:reshape':
##
##     rename

## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library("diagram")
```

```
## Warning: package 'diagram' was built under R version 4.1.1
```

```
## Loading required package: shape
```

```
## Warning: package 'shape' was built under R version 4.1.1
```

(i) Generate the transition probabilities. ## Matrix

```
#H_Sons (Shows the sons of Harvard men
#Y_Sons, shows the sons of Yale men
#D_Sons, shows the sons of Dartouth men

Mat <- matrix(c(0.6, 0, 0.4, 0.35, 0.35,0.3,0.3,0.5,0.2), nrow = 3, byrow =TRUE)
rownames(Mat) <- c("H_Sons","Y_Sons", "D_Sons")
colnames(Mat) <- c("H", "D", "Y")
Mat
```

```
##          H    D   Y
## H_Sons 0.60 0.00 0.4
## Y_Sons 0.35 0.35 0.3
## D_Sons 0.30 0.50 0.2
```

(ii) Draw the transition states with their respective probabilities ## Plotting

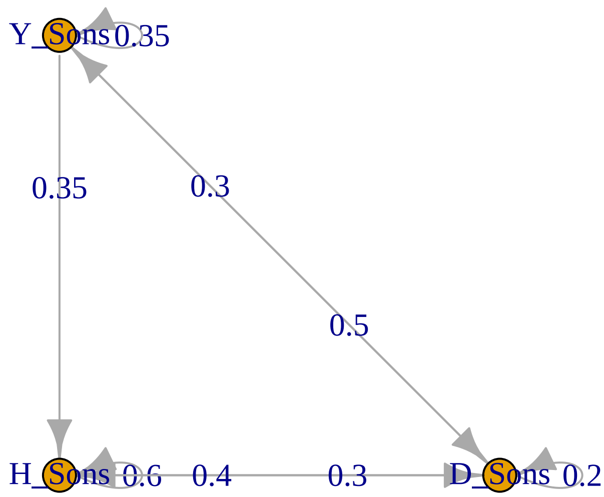
```
print(Mat)
```

```
##          H    D   Y
## H_Sons 0.60 0.00 0.4
## Y_Sons 0.35 0.35 0.3
## D_Sons 0.30 0.50 0.2
```

```
markov2 <- new('markovchain',transitionMatrix=Mat, # These are the transition probabilities of a random
               states = c('H_Sons', 'Y_Sons', 'D_Sons'))

layout <- matrix(c(0,0,0,1,1,0), ncol = 2, byrow = TRUE)

plot(markov2, nodesize = 10, layout= layout)
```



iii) Find the probability that the grandson of a man from Harvard went to Harvard.

```r
# P_grandson = (PHH_Sons*PHH_Sons)+(PHY_Sons*PY_SonsH)+(PHD_Sons*PD_SonsH)
H_grandson = (.6*.60)+(.35*.4)+(.3*0)
H_grandson
```

```
## [1] 0.5
```

iv) Modify the above by assuming that the son of a Harvard man always went to Harvard.

```r
Mat2 <- matrix(c(1,0,0,0.35,0.35, 0.3, 0.3,0.5,0.2),nrow = 3, byrow = TRUE)
rownames(Mat2) <- c("H_Sons","Y_Sons", "D_Sons")
colnames(Mat2) <- c("H", "D", "Y")
Mat2
```

```
##           H    D   Y
## H_Sons 1.00 0.00 0.0
## Y_Sons 0.35 0.35 0.3
## D_Sons 0.30 0.50 0.2
```

v) Find the probability that the grandson of a man from Harvard went to Harvard.

```r
# P_grandson = (PHH_Sons*PHH_Sons)+(PHY_Sons*PY_SonsH)+(PHD_Sons*PD_SonsH)
H_grandson = (1*1)+(.35*.0)+(.3*0)
H_grandson
```

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

vi) Determine the steady-state probabilities

```r
steadyStates(markov2)
```

```
##          H_Sons    Y_Sons   D_Sons
## [1,] 0.4457831 0.2409639 0.313253
```

Question 2 Assume that a man's profession can be classified as professional, skilled labourer, or unskilled labourer. Assume that, of the sons of professional men, 70 percent are professional, 25 percent are skilled labourers, and 5 percent are unskilled labourers. In the case of sons of skilled labourers, 58 percent are skilled labourers, 22 percent are professional, and 30 percent are unskilled. Finally, in the case of unskilled labourers, 70 percent of the sons are unskilled labourers, and 15 percent each are in the other two categories. Assume that every man has at least one son, and form a Markov chain by following the profession of a randomly chosen son of a given family through several generations.

i) Set up the matrix of trasition probabilities

```r
#P_m (Shows the sons of Professional men
#S_m, shows the sons of Skilled men
#U_m, shows the sons of Unskilled men

Prof <- matrix(c(0.7, 0.25, 0.05, 0.22, 0.58,0.2,0.15,0.15,0.7), nrow = 3, byrow =TRUE)
rownames(Prof) <- c("P_m","S_m", "U_m")
colnames(Prof) <- c("P", "S", "U")
Prof
```

```
##        P    S    U
## P_m 0.70 0.25 0.05
## S_m 0.22 0.58 0.20
## U_m 0.15 0.15 0.70
```

(ii) Draw the transition states with their respective probabilities

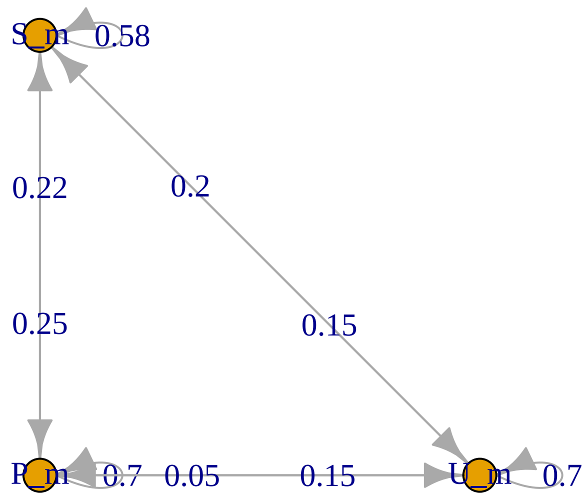
```r
print(Prof)
```

```
##        P    S    U
## P_m 0.70 0.25 0.05
## S_m 0.22 0.58 0.20
## U_m 0.15 0.15 0.70
```

```r
Prof2 <- new('markovchain',transitionMatrix=Prof, # These are the transition probabilities of a random
              states = c('P_m', 'S_m', 'U_m'))

layout <- matrix(c(0,0,0,1,1,0), ncol = 2, byrow = TRUE)

plot(Prof2, node.size = 10, layout= layout)
```



iii) Find the probability that a randomly chosen grandson of an unskilled labourer is a professional man.

```
# P_grandson = (PUU_m*PUU_m)+(PUS_m*PS_mU)+(PUP_m*PP_mU)
U_grandson = (.7*.7)+(.2*0.15)+(.05*0.15)
U_grandson
```

```
## [1] 0.5275
```

(iv) Determine the steady-state probabilities

```
steadyStates(Prof2)
```

```
##            P_m       S_m       U_m
## [1,] 0.3847695 0.3306613 0.2845691
```

Question 3 Consider the following Markov Chain

```
Mk_chain <- matrix(c(0.5, 0.25, 0.25, 0.33, 0, 0.67,0.5,0.5,0), nrow = 3, byrow =TRUE)
rownames(Mk_chain) <- c("1","2", "3")
colnames(Mat) <- c("1", "2", "3")
Mk_chain
```

```
##    [,1] [,2] [,3]
## 1 0.50 0.25 0.25
## 2 0.33 0.00 0.67
## 3 0.50 0.50 0.00
```

a) Is this chain irreducible?

Yes since every state can be reached from other state. For instance state 2 can be reached from state 1 and 3, while state 3 can be reached from state 2 and 1.

b) Is this chain aperiodic?

Yes, Meaning that none of the state is periodic or visited after a number of times. At least each state is visited at least once. There is a self transition i.e P11 > 0

c) Find the stationary distribution for this chain.

$$\pi 1 = \frac{1}{2}\pi 1 + \frac{1}{3}\pi 2 + \frac{1}{2}\pi 3$$
$$\pi 2 = \frac{1}{4}\pi 1 + \frac{1}{2}\pi 3$$
$$\pi 3 = \frac{1}{4}\pi 1 + \frac{2}{3}\pi 2$$

$$\pi 1 + \pi 2 + \pi 3 = 1$$

$$\pi 1 = 0.457, \pi 2 = 0.257, \pi 1 = 0.286$$

d) Is the stationary distribution a limiting distribution for the chain? Yes since it is irreducible and aperiodic

Hidden Markov Chains

Question 4

For each of the following data sets, is it appropriate to use HMM? Provide a one sentence explanation to your answer.

i) Stock market price data. Yes it uses HMM since Stock Price is time sensitive
ii) Appraisal of a secondary school mathematics teacher. Yes, the appraisal of the maths teacher is to understand the quality of maths and the effects on students
iii) Collaborative filtering on a database of movie reviews: For example; Netflix challenge: Predict about how much someone is going to enjoy a movie based on their and other users' movie preferences.

No it does not use HMM since User Preference don't change much over time

iv) Daily weather forecast in Nairobi

It uses HMM since the weather yesterday affects what kind of whether it will be today.

v) Optical character recognition

Yes, word recognition is sensitive to character recognition.

vi) Cost of gemstones in Bangladesh. Yes it uses the HMM since it will checks the past events which could be affected by the economy or how valuable it is in the country.

Question 5

Show that if any elements of the parameters (start probability) or A (transition probability) for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.

In the E step, since and A are initialized to be zero, there wouldn't be any training example associated with the zero probability states, nor transition to any zero probability transitions. Hence, in the M step, the updated probabilities will remain zero.