

Presentació

Us demanem en aquesta activitat que resolgueu el cas d'ús proposat mitjançant el modelat amb Cadenes de Markov. Aquest cas d'ús us permetrà posar en pràctica els conceptes treballats en aquest repte, entendre i agafar destresa en la seva aplicació a un cas d'ús concret utilitzant dades reals o realistes. Veureu també la necessitat d'utilitzar un llenguatge de programació com per exemple R per a la seva resolució i agafareu destresa en la seva utilització.

Competències

En aquesta pràctica es treballen les següents competències del Grau en Ciència de Dades Aplicada:

- Que els estudiants hagin demostrat tenir i comprendre coneixements en un àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es recolza en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda del seu camp d'estudi.
- Utilitzar de forma combinada els fonaments matemàtics, estadístics i de programació per desenvolupar solucions a problemes en l'àmbit de la ciència de les dades.
- Ús i aplicació de les TIC en l'àmbit acadèmic i professional.

Objectius

Els objectius concrets d'aquesta Pràctica són:

- Comprendre la utilitat dels conceptes d'àlgebra lineal que s'han treballat en els reptes 1-3 en l'aplicació en l'àmbit de la ciència de dades mitjançant els models matricials.
- Ser capaç de resoldre un problema utilitzant models matricials en un cas d'ús utilitzant dades reals o realistes.
- Entendre la utilitat d'utilitzar un llenguatge de programació pel tractament de grans volums de dades.
- Agafar destresa en la utilització del llenguatge R per a la resolució de problemes amb un gran volum de dades.

Descripció de la Pràctica

Saber modelitzar un sistema mitjançant sistemes dinàmics a temps discret ens pot ser molt útil en la nostra carrera professional en l'àmbit de la ciència de dades. L'exemple més famós d'aquest tipus de models és el de l'agorisme PageRank que utilitza el cercador Google.

Per una banda, us demanem que respongueu un **qüestionari** (el podeu trobar a l'aula Moodle entrant a l'enllaç "Qüestionaris" a la part dreta de l'aula) en el què treballarem la part més instrumental d'aquest repte en una sèrie de preguntes genèriques.

Us demanem també que resolgueu la pràctica descrita en aquest document. Aquests exercicis us plantejaran escenaris propis de la ciència de dades i veureu com els conceptes treballats en aquest repte tenen rellevància en aquests contextos.

Recursos

Recursos Bàsics

- Capítol *Cadenas de Markov de tiempo discreto* del llibre *Una introducción amable a la teoría de colas*.
- *Introducción a la teoría de matrices positivas: aplicaciones*.
- Document Problemes sobre models matricials en la ciència de dades.

Recursos Complementaris

- Cas d'ús i guia de resolució en R.

Criteris d'avaluació

- La pràctica s'ha de resoldre de manera individual.
- És necessari justificar tots els passos realitzats a la resolució de la Pràctica.

Tingueu en compte que les dues activitats que es plantegen en aquest repte (la resolució de la pràctica que es planteja en aquest document i el qüestionari) seran part de la nota de pràctiques ($Pr = (Pr1 + Pr2)/2$). La nota d'aquestes activitats correspon a la Pr2 (amb un pes del 20% pel qüestionari i un 80% per a la pràctica). Per a més informació sobre el model d'avaluació de l'assignatura, consulteu el pla docent.

Format i data de lliurament

Cal lliurar un únic document en PDF que incorpori la resolució de la pràctica (memòria tècnica detallada), el codi R i les imatges o figures que se us demanen.

Com s'ha dit en l'apartat anterior, s'ha de lliurar tot en un únic fitxer PDF que tingui per nom `Pr1Cognom1Cognom2Nom.pdf`, tot i que després el Registre d'Avaluació Continuada (RAC) canviarà el nom del fitxer per incloure-hi el dia i l'hora del lliurament. Aquest fitxer s'ha de lliurar en l'espai del registre de l'avaluació continuada (RAC) de l'aula abans de les 24:00 hores del dia 18/06/2021 (hora central europea d'estiu (CEST)). No s'acceptaran lliuraments fora de termini.

1 Cadenes de Markov discretes / El ratolí i el laberint

Es col·loca un ratolí al laberint que es mostra a continuació i, cada minut, es mou d'una habitació a una altra a l'atzar. Un cop el ratolí està en una habitació, el ratolí escollirà una porta que el portarà a l'habitació següent amb probabilitats idèntiques. Un cop el ratolí arriba a l'habitació 1, troba menjar (un fantàstic tros d'emmental de la vall suïssa de l'Emme al cantó de Berna) i mai no surt d'aquesta habitació. De la mateixa manera, quan arriba a l'habitació 5, queda atrapat i no pot sortir d'aquesta habitació.

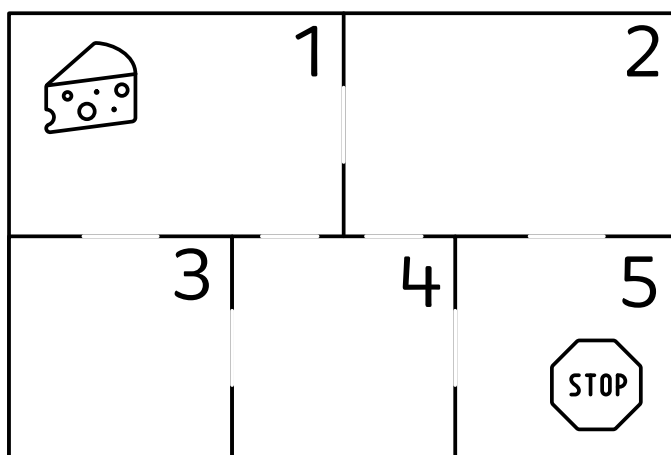


Figura 1: El ratolí i el laberint.

Taula 1: Matriu de transició entre la ubicació actual del ratolí i la ubicació del ratolí al cap d'un minut.

Ubicació del ratolí	Ubicació al cap d'un minut				
	1	2	3	4	5
1					
2					
3					
4					
5					

1. [10%] Completeu la matriu de transició de la Taula 1 tot seguint les indicacions descrites en l'enunciat de la pràctica.
2. [5%] Considereu la matriu de transició de la Taula 1 (ja completada) i definiu la matriu de transició P per files (no per columnes!). Si genereu un vector x amb els 25 valors de la taula (per files), podeu generar la matriu P de la següent manera:

```

1 labels<-c("1","2","3","4","5")
2 byRow <- TRUE
3 P<-matrix(data=x,byrow=byRow,nrow=5,dimnames=list(labels,labels))

```

3. [5%] Comproveu que la suma de les probabilitats de cada una de les cinc files és 1.
4. [5%] Creeu la cadena de Markov discreta definida per la matriu de transicions P , i anomeu-la mcP . Ajudeu-vos del següent codi, on P és la matriu definida en la pregunta 2:

```

1 install.packages("markovchain")
2 library("markovchain")
3 mcP<-new("markovchain",states=labels,byrow=byRow,transitionMatrix=P,name="The
  mouse and the maze")

```

5. [5%] Feu un diagrama de la vostra cadena de Markov discreta. Podeu ajudar-vos del següent codi:

```

1 plot(mcP,package="diagram",cex=.6)

```

6. [10%] Una cadena de Markov es diu *regular* (també *primitiva* o *ergòdica*) si hi ha alguna potència positiva de la matriu de transició les entrades de la qual siguin totes estrictament majors que zero. Però no serà necessari mirar les potències de la matriu de transició mcP^m per a tot m fins a l'infinit (afortunadament!). Un teorema ens diu que és suficient examinar les potències mcP^m per a valors de m més petits o iguals a $(n-1)^2 + 1$ on n és el nombre d'estats (en el nostre cas, $n = 5$). Genereu un programa o escriviu unes línies de codi que, donada la cadena de Markov mcP , ens retorni una llista mcL de $m = (5-1)^2 + 1 = 17$ elements, on l'element i -èssim de la llista sigui la suma del nombre de zeros de la potència mcP^i . Com a ajuda, penseu que el primer element de la llista ha de ser 14.
7. [10%] Com hem dit, la cadena de Markov serà regular si algun element de la llista mcL és zero (el que implica que alguna potència de la matriu de transició té tots els elements estrictament positius). La cadena de Markov és regular? En cas afirmatiu, tindrà un únic estat estacionari. En cas contrari, no ho podrem afirmar. Mireu si el resultat que obteniu en aquesta pregunta és compatible amb la vostra resposta de la pregunta 10.
8. [10%] La cadena de Markov mcP té dos estats absorbents. Quins són? Diries que l'existència d'estats absorbents implica la no regularitat de la cadena de Markov? Justifiqueu clarament la vostra resposta.

9. [5%] Si l'habitació a la qual es desplaça el ratolí és la tercera, que ens diu la matriu de transició de la cadena de Markov mcP sobre l'estat actual del ratolí? Justifiqueu la vostra resposta.
10. [10%] Sense fer cap tipus de càlcul, digueu i justifiqueu quantes distribucions estacionàries té la cadena de Markov mcP . Comproveu el vostre resultat amb R . També amb R , calculeu quants vectors propis de la matriu P tenen valor propi 1?
11. [10%] Aneu al qüestionari associat a la pràctica i mireu quin és el valor de N que us ha estat assignat. Quina és la probabilitat que el ratolí acabi a l'habitació 1 si ha estat inicialment ubicat a l'habitació N ?
12. [10%] I quina és la probabilitat que el ratolí acabi a l'habitació 5 si ha estat inicialment ubicat a l'habitació N ?
13. [5%] Per facilitar la correcció i la localització dels possibles errors de la programació del vostre codi en R , completeu la taula que trobareu al **qüestionari REpte 5 - Taula resum de la Pràctica 2** en relació amb les variables creades durant la resolució d'aquesta pràctica (per al vostre valor de N del mateix qüestionari):

	valor
<code>apply(P,1,sum)</code>	
<code>apply(P,2,sum)</code>	
<code>sum(mcL)</code> (pregunta 6)	
<code>sum(mcL == 0)</code> (pregunta 7)	
Nombre d'estats absorbents (pregunta 8)	
La cadena de Markov és regular? (pregunta 8)	
Des de quina habitació es desplaça el ratolí? (pregunta 9)	
Quants valors propis de P són iguals a 1? (pregunta 10)	
Probabilitat que el ratolí acabi a l'habitació 1 (pregunta 11)	
Probabilitat que el ratolí acabi a l'habitació 5 (pregunta 12)	