Aplicacions lineals, diagonalització i vectors propis

Problemes per a la ciència de dades

PID_00262382

Francesc Pozo Montero
Jordi Ripoll Missé

Universitat Oberta de Catalunya

Francesc Pozo Montero

Llicenciat en Matemàtiques per la Universitat de Barcelona (2000) i doctor en Matemàtica Aplicada per la Universitat Politècnica de Catalunya (2005). Ha estat professor associat a la Universitat Autònoma de Barcelona i professor associat, col·laborador i actualment professor agregat a la Universitat Politècnica de Catalunya. A més, és cofundador del Grup d'Innovació Matemàtica E-learning (GIMEL), responsable de diversos projectes d'innovació docent i autor de diverses publicacions. Com a membre del grup de recerca consolidat CoDAlab, centra la recerca en la teoria de control i les aplicacions en enginyeria mecànica i civil, com també en l'ús de la ciència de dades per al monitoratge de la integritat estructural i per al monitoratge de la condició, sobretot en turbines eòliques.

Jordi Ripoll Missé

Llicenciat en Matemàtiques i doctor en Ciències Matemàtiques per la Universitat de Barcelona (2005). Professor col·laborador de la Universitat Oberta de Catalunya des del 2011 i professor del Departament d'Informàtica, Matemàtica Aplicada i Estadística de la Universitat de Girona (UdG) des del 1996, on actualment és professor agregat i desenvolupa tasques de recerca en l'àmbit de la biologia matemàtica (models amb equacions en derivades parcials i dinàmica evolutiva). També ha estat professor i tutor de la UNED en dues etapes, primer al centre associat de Terrassa i actualment al de Girona. Ha participat en nombrosos projectes d'innovació docent, especialment pel que fa a l'aprenentatge de les matemàtiques en línia.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Cristina Cano Bastidas (2019)

Primera edició: febrer 2019

Av. Tibidabo, 39-43, 08035 Barcelona
Disseny: Manel Andreu
Realització editorial: Oberta UOC Publishing, SL

Índex

1. Aplicació a l'estudi de sistemes dinàmics. Estudi d'un cas

Enunciat del cas

El total de línies mòbils està repartit bàsicament entre tres companyies: Movistar, Orange i Vodafone. Mes rere mes, a causa de la competència i les ofertes agressives, es produeixen fugues de clients i captació per part de les companyies de la competència. En particular:

- Movistar és capaç de capturar el 20% del total de clients d'Orange i també el 20% del total de clients de Vodafone.
- Orange és capaç de capturar el 20% del total de clients de Movistar, però no capta cap client de Vodafone.
- Finalment, Vodafone és capaç de capturar el 30% del total de clients de Movistar i el 40% dels clients d'Orange.

Suposant que el nombre total de línies mòbils es manté constant, es demana el següent:

1) Plantegeu el sistema d'equacions que representa la distribució de línies mòbils.

Nota: si $x(t)$,$y(t)$ i $z(t)$ representen, respectivament, el nombre de línies mòbils el mes t de Movistar, Orange i Vodafone, es tractarà d'obtenir —a partir de l'esquema de distribució anterior— un sistema d'equacions de la forma:

$$x(t+1) = a_{11}x(t) + a_{12}y(t) + a_{13}z(t)$$

$$y(t+1) = a_{21}x(t) + a_{22}y(t) + a_{23}z(t)$$

$$z(t+1) = a_{31}x(t) + a_{32}y(t) + a_{33}z(t)$$

on a_{ij}, $i,j = 1,2,3$ són els coeficients que s'han de determinar en funció de les dades del problema.

2) Escriviu el sistema anterior en forma matricial, essent M la matriu de coeficients.

3) Denotant per 0 el mes inicial i per k un mes qualsevol, raoneu breument per què es complirà que:

$$\begin{pmatrix} x(k) \\ y(k) \\ z(k) \end{pmatrix} = \mathbf{M}^k \begin{pmatrix} x(0) \\ y(0) \\ z(0) \end{pmatrix}$$

Observació: l'expressió anterior ens permet predir el nombre de línies mòbils que tindrà cada companyia el mes k a partir del nombre de línies mòbils inicials.

4) Trobeu els valors propis (VAP) i els vectors propis (VEP) de $\mathbf{M}$, i també la seva matriu diagonal associada.

5) Sabent que el nombre inicial de línies mòbils és $x(0) = 350$, $y(0) = 500$ i $z(0) = 200$ (en desenes de milers), utilitzeu els resultats que s'han obtingut en l'apartat anterior per a predir el nombre de línies mòbils de cada companyia després de $k = 1$, $k = 2$ i $k = 100$ mesos.

Resolució del cas

1) El sistema d'equacions és:

$$x(t+1) = 0.5x(t) + 0.2y(t) + 0.2z(t)$$

$$y(t+1) = 0.2x(t) + 0.4y(t)$$

$$z(t+1) = 0.3x(t) + 0.4y(t) + 0.8z(t)$$

Cal destacar que si sumem els coeficients que acompanyen a $x(t)$ el resultat és 1. El mateix passa si sumem els coeficients que acompanyen a $y(t)$ i a $z(t)$, respectivament. Altrament, si alguna de les sumes d'aquests coeficients fos inferior a 1, significaria que el nombre de clients disminueix.

2) Matricialment:

$$\mathbf{M} = \begin{pmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0 \\ 0.3 & 0.4 & 0.8 \end{pmatrix}$$

Aleshores, el sistema es pot escriure com:

$$\begin{pmatrix} x(t+1) \\ y(t+1) \\ z(t+1) \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$

3) Observeu que:

$$\begin{pmatrix} x(k) \\ y(k) \\ z(k) \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} x(k-1) \\ y(k-1) \\ z(k-1) \end{pmatrix} = \mathbf{M} \cdot \mathbf{M} \cdot \begin{pmatrix} x(k-2) \\ y(k-2) \\ z(k-2) \end{pmatrix}$$

$$= \mathbf{M} \cdot \mathbf{M} \cdot \mathbf{M} \cdot \begin{pmatrix} x(k-3) \\ y(k-3) \\ z(k-3) \end{pmatrix} = \cdots$$

$$= \underbrace{\mathbf{M} \cdot \mathbf{M} \cdots \mathbf{M}}_{k \text{ vegades}} \cdot \begin{pmatrix} x(k-k) \\ y(k-k) \\ z(k-k) \end{pmatrix} = \mathbf{M}^k \begin{pmatrix} x(0) \\ y(0) \\ z(0) \end{pmatrix}$$

4) Valors i vectors propis trobats amb el programari R:

Figura 1

```
> a<-c(0.5,0.2,0.2,0.2,0.4,0,0.3,0.4,0.8)
> A<-matrix(a,nrow=3,ncol=3,byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]  0.5  0.2  0.2
[2,]  0.2  0.4  0.0
[3,]  0.3  0.4  0.8
> r<-eigen(A)
> r$values
[1] 1.0 0.4 0.3
> r$vectors
          [,1]          [,2]       [,3]
[1,] 0.4150287  2.055734e-16  0.4082483
[2,] 0.1383429 -7.071068e-01 -0.8164966
[3,] 0.8992288  7.071068e-01  0.4082483
> d<-c(1.0,0,0,0,0.4,0,0,0,0.3)
> D<-matrix(d,nrow=3,ncol=3,byrow=TRUE)
> D
     [,1] [,2] [,3]
[1,]    1  0.0  0.0
[2,]    0  0.4  0.0
[3,]    0  0.0  0.3
> P<-r$vectors
> P
          [,1]          [,2]       [,3]
[1,] 0.4150287  2.055734e-16  0.4082483
[2,] 0.1383429 -7.071068e-01 -0.8164966
[3,] 0.8992288  7.071068e-01  0.4082483
```

Noteu que les columnes de la matriu **P** representen els vectors propis. En aquest cas, han estat normalitzats, és a dir, la norma dels vectors propis és 1.

5) Per la teoria, sabem que

$$\mathbf{M}^k = \mathbf{P} \cdot \mathbf{D}^k \cdot \mathbf{P}^{-1}.$$

Per tant, fent servir R en els càlculs:

$$\text{línies}(0) = \begin{pmatrix} 350 \\ 500 \\ 200 \end{pmatrix}, \quad \text{línies}(k) = \mathbf{P} \cdot \mathbf{D}^k \cdot \mathbf{P}^{-1} \cdot \text{línies}(0)$$

Per tant,

$$\text{línies}(1) = \begin{pmatrix} 315 \\ 270 \\ 465 \end{pmatrix}, \quad \text{línies}(2) = \begin{pmatrix} 304.5 \\ 171 \\ 574.5 \end{pmatrix}, \quad \text{línies}(100) = \begin{pmatrix} 300 \\ 100 \\ 650 \end{pmatrix}$$

Figura 2

```
> b<-c(350,500,200)
> P%*%D%*%solve(P)%*%b
     [,1]
[1,]  315
[2,]  270
[3,]  465
> P%*%D^2%*%solve(P)%*%b
      [,1]
[1,] 304.5
[2,] 171.0
[3,] 574.5
> P%*%D^100%*%solve(P)%*%b
     [,1]
[1,]  300
[2,]  100
[3,]  650
```

Si haguéssim calculat el nombre de línies al mes $k = 15$, hauríem obtingut el mateix resultat que al mes $k = 100$. Això vol dir que, a partir del mes $k = 15$, el nombre de clients de cada companyia serà estable (romandrà constant). Es pot dir també que hem arribat a un estat estacionari.

2. Exercicis d'autoavaluació

En l'actualitat, les gran companyies de distribució de vídeo en reproducció en continu que es reparteixen els clients són Netflix, HBO, Amazon Prime Video i Movistar+. Atès que aquestes companyies ofereixen un servei de subscripció mensual sense permanència, existeixen fugues de clients entre elles. En particular,

- Netflix és capaç de retenir el 25% dels seus clients, alhora que capta el 25% de HBO, Amazon Prime Video i Movistar+.

- HBO no és capaç de retenir cap dels seus clients, alhora que capta el 25% de Netflix i Movistar+ i la meitat dels clients d'Amazon Prime Video.

- Amazon Prime Video no capta cap client de Movistar+, però capta el 25% de Netflix, la meitat dels clients d'HBO i reté el 25% dels seus clients.

- Finalment, Movistar+ és capaç de retenir la meitat dels seus clients, no capta cap client d'Amazon Prime Video i capta el 25% de Netflix i HBO.

A partir de la formulació matemàtica que representa l'evolució de clients de les quatre empreses de distribució de vídeo en reproducció en continu:

a) Plantegeu el sistema d'equacions que representa la distribució de línies mòbils.

Nota: si $x(t)$, $y(t)$, $z(t)$ i $v(t)$ representen, respectivament, el nombre de clients el mes t de Netflix, HBO, Amazon Prime Video i Movistar+, es tractarà d'obtenir —a partir de l'esquema de distribució anterior— un sistema d'equacions de la forma:

$$x(t+1) = a_{11}x(t) + a_{12}y(t) + a_{13}z(t) + a_{14}v(t)$$

$$y(t+1) = a_{21}x(t) + a_{22}y(t) + a_{23}z(t) + a_{24}v(t)$$

$$z(t+1) = a_{31}x(t) + a_{32}y(t) + a_{33}z(t) + a_{34}v(t)$$

$$v(t+1) = a_{41}x(t) + a_{42}y(t) + a_{43}z(t) + a_{44}v(t)$$

on els a_{ij}, $i,j = 1,2,3,4$ són els coeficients que s'han de determinar en funció de les dades del problema.

En aquest cas, el sistema d'equacions és:

$$x(t+1) = 0.25x(t) + 0.25y(t) + 0.25z(t) + 0.25v(t)$$

$$y(t+1) = 0.25x(t) + 0.50z(t) + 0.25v(t)$$

$$z(t+1) = 0.25x(t) + 0.50y(t) + 0.25z(t)$$

$$v(t+1) = 0.25x(t) + 0.25y(t) + 0.50v(t)$$

b) Escriviu el sistema anterior en forma matricial, essent **M** la matriu de coeficients.

Matricialment:

$$\mathbf{M} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0 & 0.50 & 0.25 \\ 0.25 & 0.50 & 0.25 & 0 \\ 0.25 & 0.25 & 0 & 0.50 \end{pmatrix}$$

És a dir,

$$\begin{pmatrix} x(t+1) \\ y(t+1) \\ z(t+1) \\ v(t+1) \end{pmatrix} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0 & 0.50 & 0.25 \\ 0.25 & 0.50 & 0.25 & 0 \\ 0.25 & 0.25 & 0 & 0.50 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \\ z(t) \\ v(t) \end{pmatrix}$$

c) Indiqueu quins són els valors i vectors propis associats a la matriu del model.

Trobem els valors i vectors propis amb la funció `eigen` de R. Els valors propis (`r$values`) són:

$$\lambda_1 = 1$$

$$\lambda_2 = 4.330127 \cdot 10^{-1}$$

$$\lambda_3 = 4.440892 \cdot 10^{-16} \approx 0$$

$$\lambda_4 = -4.330127 \cdot 10^{-1} = -\lambda_2$$

Els vectors propis (`r$vectors`) són:

$$v_1 = \begin{pmatrix} -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \end{pmatrix}; \quad v_2 = \begin{pmatrix} 0 \\ 0.2113249 \\ 0.5773503 \\ -0.7886751 \end{pmatrix}; \quad v_3 = \begin{pmatrix} 0.8660254 \\ -0.2886751 \\ -0.2886751 \\ -0.2886751 \end{pmatrix}; \quad v_4 = \begin{pmatrix} 0 \\ 0.788675 \\ -0.5773503 \\ -0.2113249 \end{pmatrix}$$

Noteu que els quatre vectors propis són unitaris, és a dir, tots tenen norma 1.

Figura 3

```
> a<-c(0.25,0.25,0.25,0.25,0.25,0,0.50,0.25,0.25,0.50,0.25,0,0.25,0.25,0,0
.50)
> M<-matrix(a,nrow=4,ncol=4,byrow=TRUE)
> M
     [,1] [,2] [,3] [,4]
[1,] 0.25 0.25 0.25 0.25
[2,] 0.25 0.00 0.50 0.25
[3,] 0.25 0.50 0.25 0.00
[4,] 0.25 0.25 0.00 0.50
> r<-eigen(M)
> r$values
[1]  1.000000e+00  4.330127e-01  4.440892e-16 -4.330127e-01
> r$vectors
     [,1]       [,2]       [,3]       [,4]
[1,] -0.5  0.0000000  0.8660254  0.0000000
[2,] -0.5  0.2113249 -0.2886751  0.7886751
[3,] -0.5  0.5773503 -0.2886751 -0.5773503
[4,] -0.5 -0.7886751 -0.2886751 -0.2113249
> |
```

d) Quina és la distribució dels clients a llarg termini ($k \to +\infty$) si en l'actualitat, ($k = 0$), la proporció de clients de Netflix, HBO i Amazon és del 20% mentre que la proporció de clients de Movistar+ és del 40%?

D'una banda, sabem que la distribució dels clients el mes k és igual a:

$$\begin{pmatrix} x(k) \\ y(k) \\ z(k) \\ v(k) \end{pmatrix} = \mathbf{M}^k \begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ v(0) \end{pmatrix}$$

D'altra banda, per la teoria sabem que la matriu $\mathbf{M}^k$ es pot expressar com:

$$\mathbf{M}^k = \mathbf{P} \cdot \mathbf{D}^k \cdot \mathbf{P}^{-1}$$

on $\mathbf{P}$ és la matriu formada pels vectors propis i $\mathbf{D}$ és la matriu diagonal formada pels valors propis. Aleshores, a llarg termini, és a dir, quan k tendeix a infinit, la distribució vindrà donada per:

$$\mathbf{P} \cdot \left(\lim_{k \to +\infty} \mathbf{D}^k \right) \cdot \mathbf{P}^{-1} \cdot \begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ v(0) \end{pmatrix}$$

El càlcul del límit $\lim_{k \to +\infty} \mathbf{D}^k$ és senzill:

$$\lim_{k \to +\infty} \mathbf{D}^k = \lim_{k \to +\infty} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4.330127 \cdot 10^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4.330127 \cdot 10^{-1} \end{pmatrix}^k$$

$$= \lim_{k \to +\infty} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \left(4.330127 \cdot 10^{-1}\right)^k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(-4.330127 \cdot 10^{-1}\right)^k \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lim_{k \to +\infty} \left(4.330127 \cdot 10^{-1}\right)^k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lim_{k \to +\infty} \left(-4.330127 \cdot 10^{-1}\right)^k \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Per tant, tenim que la distribució a llarg termini és:

$$\underbrace{\begin{pmatrix} -0.5 & 0.0000000 & 0.8660254 & 0.0000000 \\ -0.5 & 0.2113249 & -0.2886751 & 0.7886751 \\ -0.5 & 0.5773503 & -0.2886751 & -0.5773503 \\ -0.5 & -0.7886751 & -0.2886751 & -0.2113249 \end{pmatrix}}_{\mathbf{P}} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \mathbf{P}^{-1} \cdot \begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ v(0) \end{pmatrix}$$

on:

$$\begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} 0.20 \\ 0.20 \\ 0.20 \\ 0.40 \end{pmatrix}$$

El resultat és, finalment:

$$\mathbf{P} \cdot \left(\lim_{k \to +\infty} \mathbf{D}^k \right) \cdot \mathbf{P}^{-1} \cdot \begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}$$

A llarg termini, la distribució de clients entre les companyies de vídeo en reproducció en continu serà de 25% per a cada companyia.