Aplicaciones lineales, diagonalización y vectores propios

Problemas para la ciencia de datos

PID_00262383

Francesc Pozo Montero
Jordi Ripoll Missé

Universitat Oberta de Catalunya

Francesc Pozo Montero

Licenciado en Matemáticas por la Universidad de Barcelona (2000) y doctor en Matemática Aplicada por la Universidad Politécnica de Cataluña (2005). Ha sido profesor asociado de la Universidad Autónoma de Barcelona y profesor asociado, colaborador y actualmente profesor agregado en la Universidad Politécnica de Cataluña. Además, es cofundador del Grupo de Innovación Matemática E-learning (GIMEL), responsable de varios proyectos de innovación docente y autor de varias publicaciones. Como miembro del grupo de investigación consolidado CoDAlab, centra su investigación en la teoría de control y las aplicaciones en ingeniería mecánica y civil, así como en el uso de la ciencia de datos para la monitorización de la integridad estructural y para la monitorización de la condición, sobre todo en turbinas eólicas.

Jordi Ripoll Missé

Licenciado en Matemáticas y doctor en Ciencias Matemáticas por la Universidad de Barcelona (2005). Profesor colaborador de la Universitat Oberta de Catalunya desde 2011 y profesor del Departamento de Informática, Matemática Aplicada y Estadística de la Universidad de Girona (UdG) desde 1996, donde actualmente es profesor agregado y desarrolla tareas de investigación en el ámbito de la biología matemática (modelos con ecuaciones en derivadas parciales y dinámica evolutiva). También ha sido profesor y tutor de la UNED en dos etapas, primero en el centro asociado de Terrassa y actualmente en el de Girona. Ha participado en numerosos proyectos de innovación docente, especialmente en cuanto al aprendizaje de las matemáticas en línea.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por la profesora: Cristina Cano Bastidas (2019)

Primera edición: febrero 2019
© Francesc Pozo Montero, Jordi Ripoll Missé
Todos los derechos reservados
© de esta edición, FUOC, 2019
Av. Tibidabo, 39-43, 08035 Barcelona
Diseño: Manel Andreu
Realización editorial: Oberta UOC Publishing, SL

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Índice general

1. Aplicación al estudio de sistemas dinámicos. Estudio de un caso

Enunciado del caso

El total de líneas móviles está repartido básicamente entre tres compañías: Movistar, Orange y Vodafone. Cada mes, debido a la competencia y las ofertas agresivas, se producen fugas de clientes y captación por parte de las compañías de la competencia. En particular:

- Movistar es capaz de capturar el 20 % del total de clientes de Orange y también el 20 % del total de clientes de Vodafone.

- Orange es capaz de capturar el 20 % del total de clientes de Movistar, pero no capta ningún cliente de Vodafone.

- Finalmente, Vodafone es capaz de capturar el 30 % del total de clientes de Movistar y el 40 % de los clientes de Orange.

Suponiendo que el número total de líneas móviles se mantiene constante, se pide lo siguiente:

1) Plantead el sistema de ecuaciones que representa la distribución de líneas móviles.

Nota: si $x(t)$,$y(t)$ y $z(t)$ representan, respectivamente, el número de líneas móviles el mes t de Movistar, Orange y Vodafone, habrá que obtener —a partir del esquema de distribución anterior— un sistema de ecuaciones de la forma:

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

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

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

donde a_{ij}, i,j = 1,2,3 son los coeficientes que se tienen que determinar en función de los datos del problema.

2) Escribid el sistema anterior en forma matricial, siendo M la matriz de coeficientes.

3) Dado un mes inicial 0 y un mes cualquiera, k, razonad brevemente por qué se cumplirá que:

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

Observación: la expresión anterior nos permite predecir el número de líneas móviles que tendrá cada compañía el mes k a partir del número de líneas móviles iniciales.

4) Encontrad los valores propios (VAP) y los vectores propios (VEP) de $\mathbf{M}$, y también la matriz diagonal asociada a estos.

5) Dado el siguiente número inicial de líneas móviles: $x(0) = 350$, $y(0) = 500$ y $z(0) = 200$ (en decenas de miles), utilizad los resultados que se han obtenido en el apartado anterior para predecir el número de líneas móviles de cada compañía después de $k = 1$, $k = 2$ y $k = 100$ meses.

Resolución del caso

1) El sistema de ecuaciones es:

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

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

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

Hay que destacar que si sumamos los coeficientes que acompañan a $x(t)$ el resultado es 1. Lo mismo ocurre si sumamos los coeficientes que acompañan a $y(t)$ y a $z(t)$, respectivamente. De lo contrario, si alguna de las sumas de estos coeficientes fuera inferior a 1, significaría que el número de clientes disminuye.

2) Matricialmente:

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

Entonces, el sistema se puede escribir así:

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

3) Observad que:

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

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

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

4) Valores y vectores propios calculados con el software 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
```

Observad que las columnas de la matriz **P** representan los vectores propios. En este caso, han sido normalizados, es decir, la norma de los vectores propios es 1.

5) Por la teoría, sabemos que

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

Por lo tanto, usando R en los cálculos:

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

Por lo tanto,

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

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 hubiéramos calculado el número de líneas el mes $k = 15$, habríamos obtenido el mismo resultado que en el mes $k = 100$. Esto quiere decir que, a partir del mes $k = 15$, el número de clientes de cada compañía será estable (permanecerá constante). Se puede decir también que hemos llegado a un estado estacionario.

2. Ejercicios de autoevaluación

En la actualidad, las grandes compañías de distribución de vídeo en emisión en línea que se reparten los clientes son Netflix, HBO, Amazon Prime Video y Movistar+. Dado que estas compañías ofrecen un servicio de suscripción mensual sin permanencia, existen fugas de clientes entre ellas. En particular,

- Netflix es capaz de retener el 25 % de sus clientes, a la vez que capta el 25 % de HBO, Amazon Prime Video y Movistar+.

- HBO no es capaz de retener ninguno de sus clientes, a la vez que capta el 25 % de Netflix y Movistar+ y la mitad de los clientes de Amazon Prime Video.

- Amazon Prime Video no capta ningún cliente de Movistar+, pero capta el 25 % de Netflix, la mitad de los clientes de HBO y retiene el 25 % de sus clientes.

- Finalmente, Movistar+ es capaz de retener la mitad de sus clientes, no capta ningún cliente de Amazon Prime Video y capta el 25 % de Netflix y HBO.

A partir de la formulación matemática que representa la evolución de clientes de las cuatro empresas de distribución de vídeo en emisión en línea:

a) Plantead el sistema de ecuaciones que representa la distribución de líneas móviles.

Nota: si $x(t)$, $y(t)$, $z(t)$ y $v(t)$ representan, respectivamente, el número de clientes el mes t de Netflix, HBO, Amazon Prime Video y Movistar+, habrá que obtener —a partir del esquema de distribución anterior— un sistema de ecuaciones de la forma:

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

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

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

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

donde a_{ij}, $i,j = 1,2,3,4$ son los coeficientes que se tienen que determinar en función de los datos del problema.

En este caso, el sistema de ecuaciones es:

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

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

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

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

b) Escribid el sistema anterior en forma matricial, siendo $\mathbf{M}$ la matriz de coeficientes.

Matricialmente:

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

Es decir,

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

c) Indicad cuáles son los valores y vectores propios asociados a la matriz del modelo.

Encontramos los valores y vectores propios con la función `eigen` de R. Los valores propios (`r$values`) son:

$$\lambda_1 = 1$$

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

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

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

Los vectores propios (`r$vectores`) son:

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

Observad que los cuatro vectores propios son unitarios, es decir, todos tienen 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) Cuál es la distribución de los clientes a largo plazo ($k \to +\infty$) si en la actualidad, ($k = 0$), la proporción de clientes de Netflix, HBO y Amazon es del 20 % mientras que la proporción de clientes de Movistar+ es del 40 %?

Por un lado, sabemos que la distribución de los clientes el mes k es igual a:

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

Por otro lado, por la teoría sabemos que la matriz $\mathbf{M}^k$ se puede expresar como:

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

donde $\mathbf{P}$ es la matriz formada por los vectores propios y $\mathbf{D}$ es la matriz diagonal formada por los valores propios. Entonces, a largo plazo, es decir, cuando k tiende a infinito, la distribución vendrá dada por:

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

El cálculo del límite $\lim_{k \to +\infty} \mathbf{D}^k$ es sencillo:

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

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

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

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

Por lo tanto, tenemos que la distribución a largo plazo es:

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

donde:

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

El resultado es, finalmente:

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

A largo plazo, la distribución de clientes entre las compañías de vídeo en emisión en línea será de 25 % para cada compañía.