
Descomposició en valores singulares: introducció y aplicaciones

Análisis de componentes principales (PCA) y descomposició en valores singulares (SVD)

PID_00262387

Francesc Pozo Montero
Núria Parés Mariné

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.

Núria Parés Maríné

Licenciada en Matemáticas por la Universidad Politécnica de Cataluña (1999) y doctora en Matemática Aplicada por la Universidad Politécnica de Cataluña (2005). Es profesora de la Universidad Politécnica de Cataluña desde el año 2000 –actualmente, como profesora agregada–, cofundadora del Grupo de Innovación Matemática E-learning (GIMEL), responsable de diferentes proyectos de innovación docente y autora de varias publicaciones y de libros docentes. Como miembro del grupo consolidado LaCàN (UPC), centra su investigación en el desarrollo de técnicas eficientes para la resolución numérica de ecuaciones en derivadas parciales y en la estimación del error asociado a estas simulaciones numéricas.

La revisión de este recurso de aprendizaje UOC ha sido coordinada por la profesora: Cristina Cano Bastidas

Segunda edición: septiembre 2020
© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)
Av. Tibidabo, 39-43, 08035 Barcelona
Autoría: Francesc Pozo Montero, Núria Parés Maríné
Producción: FUOC
Todos los derechos reservados

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 este eléctrico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita del titular de los derechos.

Índice general

1	La maldición de la dimensión	5
1.1	Ejemplo introductorio: la interpolación polinómica pura	5
1.2	Otro ejemplo: la encuesta de presupuestos familiares	7
2	Análisis de componentes principales	9
2.1	Preprocesamiento: el escalado de los datos	11
2.2	Matriz de covarianzas	14
2.3	Diagonalización de la matriz de covarianzas	16
2.4	La matriz de covarianzas de los datos transformados	17
2.4.1	¿Cómo debemos interpretar, por ejemplo, la primera componente principal?	18
2.4.2	¿Cuál es el peso de la primera componente principal?	20
2.5	Reducción de la dimensión	22
2.5.1	El error residual	23
2.6	Ejemplo de aplicación: encuesta de presupuestos familiares ..	24
3	Descomposición en valores singulares	28
3.1	Ejemplo introductorio	28
3.2	Descomposición en valores singulares reducida	29
3.2.1	Cálculo de los valores singulares y de los vectores sin- gulares	30
3.3	Descomposición en valores singulares completa	34
3.3.1	Propiedades interesantes de la descomposición en valores singulares	38
3.4	Aplicación de la descomposición en valores singulares: compresión de imágenes	38
	Resumen	46
	Ejercicios de autoevaluación	48
	Solucionario	50
	Glosario	57
	Bibliografía	58

1. La maldición de la dimensión

1.1 Ejemplo introductorio: la interpolación polinómica pura

Un problema de interpolación clásico es encontrar, por ejemplo, la parábola que se ajusta mejor a tres puntos dados, por ejemplo:

$$(x_0, y_0), (x_1, y_1) \text{ y } (x_2, y_2).$$

La ecuación de una parábola es

$$y = ax^2 + bx + c,$$

donde a , b y c son números reales. Así pues, se trata de encontrar el valor de tres parámetros. Para hacerlo, imponemos que los tres puntos dados estén sobre la parábola. Es decir, los tres puntos tienen que satisfacer la ecuación de la parábola. Esto equivale al sistema de ecuaciones lineales siguiente:

$$y_0 = ax_0^2 + bx_0 + c$$

$$y_1 = ax_1^2 + bx_1 + c$$

$$y_2 = ax_2^2 + bx_2 + c$$

que se puede expresar en forma matricial:

$$\underbrace{\begin{bmatrix} x_0^2 & x_0 & 1 \\ x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \end{bmatrix}}_M \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

Por lo tanto, la solución del sistema —suponiendo que el determinante de la matriz M no es cero— se puede calcular así:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = M^{-1} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

Matriz de Vandermonde

Por la forma de la matriz M , en la que en cada fila se presentan los términos de una progresión geométrica, diremos que la matriz es de Vandermonde. El nombre proviene del matemático francés Alexandre-Théophile Vandermonde.

En este caso, el cálculo de la inversa de la matriz M —con tres filas y tres columnas— es asequible y lo podríamos hacer fácilmente. Pero imaginad qué pasaría si, en vez de buscar la parábola que se ajusta mejor a tres puntos dados, buscáramos el polinomio $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ de grado n que se ajusta mejor a puntos $n + 1$ dados, por ejemplo:

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$$

En este caso, el sistema de ecuaciones que tendríamos que resolver, en forma matricial, sería:

$$\underbrace{\begin{bmatrix} x_0^n & \cdots & x_0 & 1 \\ x_1^n & \cdots & x_1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_n^n & \cdots & x_n & 1 \end{bmatrix}}_{M_{n+1}} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

También en este caso, la solución del sistema —suponiendo que el determinante de la matriz M_{n+1} no es cero— se puede calcular así:

$$\begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{bmatrix} = M_{n+1}^{-1} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Ahora bien, el cálculo de la inversa de la matriz M_{n+1} ya no resulta sencillo si n es grande. Se puede recurrir, por supuesto, a métodos numéricos, pero se puede demostrar que la matriz de Vandermonde, como por ejemplo M_{n+1} , está mal condicionada. Por lo tanto, los pequeños errores numéricos que se puedan producir en el proceso de cálculo de la matriz inversa pueden afectar muy negativamente a la solución del sistema. De hecho, la solución final propuesta por el método numérico puede alejarse de la solución real.

Condicionamiento de una matriz

El número de condición de una matriz indica, por ejemplo, cómo el determinante de la matriz se puede ver afectado por pequeños cambios en los elementos que la forman. Si la matriz está mal condicionada, el sistema de ecuaciones lineales también lo estará.

En matemáticas, y más concretamente en el campo de la interpolación polinómica pura, el problema expuesto anteriormente se resuelve, por ejemplo, con el método de las diferencias divididas de Newton o con los polinomios de Lagrange. Por medio de estos dos métodos, los coeficientes de los polinomios interpoladores se calculan sin necesidad de resolver un sistema de ecuaciones lineales, lo que evita el problema del mal acondicionamiento y, por lo tanto, la maldición de la dimensionalidad.

1.2 Otro ejemplo: la encuesta de presupuestos familiares

La encuesta de presupuestos familiares (EPF) suministra información anual sobre la naturaleza y el destino de los gastos de consumo, además de varias características relativas a las condiciones de vida de los hogares. Los gastos de consumo se refieren tanto al flujo monetario que destina el hogar a pagar determinados bienes y servicios de consumo final, como al valor de los bienes percibidos en concepto de autoconsumo, autosuministro, salario en especie, etc. El tamaño de muestra es de 24.000 hogares por año, aproximadamente.

Encuesta de presupuestos familiares

El Instituto Nacional de Estadística (www.ine.es) publica anualmente la encuesta de presupuestos familiares (EPF).

La información de la encuesta se presenta de varias maneras. Por ejemplo, puede estar agrupada por comunidades autónomas —incluyendo las ciudades autónomas de Ceuta y Melilla:

- 1) Andalucía
- 2) Aragón
- 3) Asturias
- 4) Baleares
- 5) Canarias
- 6) Cantabria
- 7) Castilla y León
- 8) Castilla-La Mancha
- 9) Cataluña
- 10) Comunidad Valenciana
- 11) Extremadura
- 12) Galicia
- 13) Comunidad de Madrid
- 14) Murcia
- 15) Navarra
- 16) País Vasco
- 17) La Rioja
- 18) Ceuta
- 19) Melilla

Se miden un total de doce variables:

- 1) Alimentos y bebidas no alcohólicas
- 2) Bebidas alcohólicas y tabaco
- 3) Vestido y calzado
- 4) Vivienda, agua, electricidad, gas y otros combustibles
- 5) Muebles, artículos del hogar y artículos para el mantenimiento corriente del hogar
- 6) Sanidad
- 7) Transporte
- 8) Comunicaciones
- 9) Ocio y cultura
- 10) Enseñanza
- 11) Restaurantes y hoteles
- 12) Otros bienes y servicios

Por ejemplo, la tabla 1 muestra el gasto medio por persona (en euros) en cada una de las diecisiete comunidades autónomas y las dos ciudades autónomas españolas, en lo que respecta a alimentos y bebidas no alcohólicas. Con esta tabla de doble entrada es fácil sacar alguna conclusión. Por ejemplo, en el País Vasco el gasto medio por persona —de alimentos y bebidas no alcohólicas— es superior; en cambio, en Ceuta es donde este gasto medio es inferior. Si añadimos más columnas a la tabla 1, es decir, si incluimos la información de más variables, posiblemente las conclusiones que se podrán sacar serán más interesantes. Aun así, también es más difícil ver o inferir alguna conclusión, puesto que tendremos una matriz de datos con $12 \times 19 = 228$ gastos medios.

¿Y si en lugar de tener la información agrupada por comunidades autónomas la tuviéramos por provincias? Entonces dispondríamos de $50 \times 12 = 600$ gastos medios. Imaginaos la dificultad de obtener alguna conclusión tan solo observando esta información.

Como se puede ver, es fácil que al aumentar el número de información disponible la información resultante sea difícil de interpretar y de visualizar y, también, que sea difícil extraer alguna conclusión.

Este efecto es lo que denominamos *maldición de la dimensionalidad*. De forma más precisa, si aumentamos la dimensión de la información que tenemos, esta pasa a ser más rica. Sin embargo, a la vez es más difícil interpretarla. En este módulo veremos dos técnicas para reducir la dimensionalidad manteniendo la riqueza de la información, y aprenderemos cómo podemos expresar nuestros datos de forma que sea más fácil extraer sus características.

Provincias

Hemos tenido en cuenta cincuenta provincias españolas, es decir, no hemos considerado las dos ciudades autónomas de Ceuta y Melilla.

Tabla 1. Gasto medio por persona (en euros)

Comunidad autónoma	Alimentos y bebidas no alcohólicas
Andalucía	1.533,39
Aragón	1.755,21
Asturias	1.777,14
Baleares	1.697,69
Canarias	1.460,25
Cantabria	1.793,89
Castilla y León	1.780,34
Castilla-La Mancha	1.432,40
Cataluña	1.833,10
Comunidad Valenciana	1.513,25
Extremadura	1.317,60
Galicia	1.678,53
Comunidad de Madrid	1.639,72
Murcia	1.662,63
Navarra	1.853,64
País Vasco	1.959,03
La Rioja	1.679,08
Ceuta	1.327,42
Melilla	1.473,79

2. Análisis de componentes principales

El análisis de componentes principales (PCA, del inglés *principal component analysis*) es un método simple y no paramétrico para extraer información relevante de conjuntos de datos que pueden ser confusos. Además, puede servir para dar argumentos o indicaciones sobre cómo reducir la dimensión de un conjunto complejo de datos y así revelar posibles estructuras escondidas o características interesantes. Las aplicaciones son diversas.

Consideramos un primer ejemplo sencillo para ilustrar cómo funciona esta técnica. Se trata de un famoso conjunto de datos, el llamado *Iris*, que proporciona las medidas en centímetros de las variables de longitud y anchura del sépalo y de longitud y anchura del pétalo, respectivamente, para cincuenta flores de cada una de las especies *Iris setosa*, *Iris versicolor* e *Iris virginica*. De entrada, para simplificar el estudio de la técnica, veremos una muestra de quince flores, cinco de cada tipo, como se expone en la tabla 2.

Iris

Los datos fueron recogidos por Edgar Anderson en el año 1935 y publicados en el artículo "The irises of the Gaspé Peninsula", *Bulletin of the American Iris Society*, 59, págs. 2-5.

Tabla 2. Longitud y anchura del sépalo y longitud y anchura del pétalo (en centímetros)

flor	long. sépalo	anch. sépalo	long. pétalo	anch. pétalo	tipo
1	5.1	3.5	1.4	0.2	<i>setosa</i>
2	4.9	3.0	1.4	0.2	<i>setosa</i>
3	4.7	3.2	1.3	0.2	<i>setosa</i>
4	4.6	3.1	1.5	0.2	<i>setosa</i>
5	5.0	3.6	1.4	0.2	<i>setosa</i>
6	7.0	3.2	4.7	1.4	<i>versicolor</i>
7	6.4	3.2	4.5	1.5	<i>versicolor</i>
8	6.9	3.1	4.9	1.5	<i>versicolor</i>
9	5.5	2.3	4.0	1.3	<i>versicolor</i>
10	6.5	2.8	4.6	1.5	<i>versicolor</i>
11	6.3	3.3	6.0	2.5	<i>virginica</i>
12	5.8	2.7	5.1	1.9	<i>virginica</i>
13	7.1	3.0	5.9	2.1	<i>virginica</i>
14	6.3	2.9	5.6	1.8	<i>virginica</i>
15	6.5	3.0	5.8	2.2	<i>virginica</i>

Fuente: Edgar Anderson (1935). "The irises of the Gaspé Peninsula"

En este caso, hemos considerado una muestra de quince flores, de las cuales hemos medido cuatro variables:

- 1) longitud del sépalo
- 2) anchura del sépalo
- 3) longitud del pétalo
- 4) anchura del pétalo

En un caso general, podemos considerar que hay que medir m variables de un total de n elementos o experimentos y organizar toda esta información en una matriz \mathbf{X} de n filas y m columnas:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} \in \mathcal{M}_{n \times m}(\mathbb{R})$$

Fijaos en que $\mathcal{M}_{n \times m}(\mathbb{R})$ representa el espacio vectorial de las matrices de dimensión $n \times m$ con coeficientes reales.

De la matriz anterior, el i -ésimo vector fila

$$x_i^T = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \end{bmatrix}$$

representa los valores de todas las variables para uno de los elementos de la muestra, mientras que el j -ésimo vector columna

$$v_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{bmatrix}$$

representa el valor de la j -ésima variable para todos los elementos de la muestra.

En el caso de nuestro ejemplo,

$$x_2^T = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \end{bmatrix}$$

mientras que

$$v_3 = \begin{bmatrix} 1.4 \\ 1.4 \\ 1.3 \\ 1.5 \\ 1.4 \\ 4.7 \\ 4.5 \\ 4.9 \\ 4.0 \\ 4.6 \\ 6.0 \\ 5.1 \\ 5.9 \\ 5.6 \\ 5.8 \end{bmatrix}$$

2.1 Preprocesamiento: el escalado de los datos

Dado que las variables pueden tener diferentes escalas y magnitudes, hay que aplicar a la matriz de datos un preproceso para escalar las variables, de forma que todas tengan de media 0 y de desviación típica 1.

Denominamos μ_j a la media aritmética de la variable j -ésima, que se define así:

$$\mu_j = \frac{1}{n} \sum_{i=1}^n x_{ij} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n}$$

Del mismo modo, denominamos σ_j^2 a la varianza de la variable j -ésima, que se define así:

$$\sigma_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \mu_j)^2 = \frac{(x_{1j} - \mu_j)^2 + (x_{2j} - \mu_j)^2 + \dots + (x_{nj} - \mu_j)^2}{n-1}$$

Finalmente, cada elemento de la matriz X se normaliza de la manera siguiente:

$$\check{x}_{ij} := \frac{x_{ij} - \mu_j}{\sigma_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

Una matriz en la que las columnas tienen media cero se llama matriz centrada.

En el caso de nuestro ejemplo, la media y la desviación típica de la primera variable son:

$$\mu_1 = \frac{5.1 + 4.9 + \dots + 6.3 + 6.5}{15} = \frac{88.6}{15} = 5.906667$$

$$\sigma_1 = \sqrt{\frac{(5.1 - \mu_1)^2 + \dots + (6.5 - \mu_1)^2}{14}} = 0.8737985$$

Desviación típica

La desviación típica se define como la raíz cuadrada de la varianza.

El resto de las medias y desviaciones típicas son:

$$\mu_2 = 3.06$$

$$\sigma_2 = 0.3180296$$

$$\mu_3 = 3.873333$$

$$\sigma_3 = 1.891887$$

$$\mu_4 = 1.246667$$

$$\sigma_4 = 0.8296873$$

Por lo tanto, la tabla 2, al normalizar sus datos, se convierte en la tabla 3.

Tabla 3. Longitud y anchura del sépalo y longitud y anchura del pétalo (datos normalizados)

flor	long. sépalo	anch. sépalo	long. pétalo	anch. pétalo	tipo
1	-0.923172415	1.383518809	-1.307336408	-1.261519508	setosa
2	-1.152058138	-0.188661656	-1.307336408	-1.261519508	setosa
3	-1.38094386	0.44021053	-1.360193675	-1.261519508	setosa
4	-1.495386722	0.125774437	-1.254479141	-1.261519508	setosa
5	-1.037615276	1.697954901	-1.307336408	-1.261519508	setosa
6	1.251241951	0.44021053	0.436953409	0.18480859	versicolor
7	0.564584783	0.44021053	0.331238874	0.305335932	versicolor
8	1.13679909	0.125774437	0.542667943	0.305335932	versicolor
9	-0.46540097	-2.389714306	0.066952538	0.064281249	versicolor
10	0.679027644	-0.817533841	0.384096141	0.305335932	versicolor
11	0.450141921	0.754646623	1.124097882	1.510609348	virginica
12	-0.122072385	-1.131969934	0.648382477	0.787445298	virginica
13	1.365684812	-0.188661656	1.071240615	1.028499981	virginica
14	0.450141921	-0.503097749	0.912668813	0.666917957	virginica
15	0.679027644	-0.188661656	1.018383347	1.149027323	virginica

Fuente: Edgar Anderson (1935). "The Irises of the Gaspé Peninsula"

Los datos de la tabla 3 también se pueden expresar en forma matricial:

$$\check{X} = \begin{bmatrix} -0.923172415 & 1.383518809 & -1.307336408 & -1.261519508 \\ -1.152058138 & -0.188661656 & -1.307336408 & -1.261519508 \\ -1.38094386 & 0.44021053 & -1.360193675 & -1.261519508 \\ -1.495386722 & 0.125774437 & -1.254479141 & -1.261519508 \\ -1.037615276 & 1.697954901 & -1.307336408 & -1.261519508 \\ 1.251241951 & 0.44021053 & 0.436953409 & 0.18480859 \\ 0.564584783 & 0.44021053 & 0.331238874 & 0.305335932 \\ 1.13679909 & 0.125774437 & 0.542667943 & 0.305335932 \\ -0.46540097 & -2.389714306 & 0.066952538 & 0.064281249 \\ 0.679027644 & -0.817533841 & 0.384096141 & 0.305335932 \\ 0.450141921 & 0.754646623 & 1.124097882 & 1.510609348 \\ -0.122072385 & -1.131969934 & 0.648382477 & 0.787445298 \\ 1.365684812 & -0.188661656 & 1.071240615 & 1.028499981 \\ 0.450141921 & -0.503097749 & 0.912668813 & 0.666917957 \\ 0.679027644 & -0.188661656 & 1.018383347 & 1.149027323 \end{bmatrix}$$

Para simplificar la notación, a pesar de que \check{X} representa la matriz de datos normalizada, continuaremos hablando de la matriz X y entenderemos, en el resto del módulo, que está normalizada.

Los datos de la tabla 3 se pueden representar gráficamente de forma sencilla si generamos diagramas de dispersión bidimensionales, para cada par de variables. En este caso, esto querría decir que podemos generar un total de

$$\binom{4}{2} = \frac{4 \cdot 3}{2} = 6$$

diagramas de dispersión. Esta cantidad parece razonable. Pero ¿que pasaría con una muestra en que hayamos medido diez variables? En ese caso, tendríamos que representar gráficamente $\binom{10}{2} = 45$ diagramas de dispersión, y nadie nos podría garantizar que alguno de estos cuarenta y cinco diagramas marcara alguna tendencia o mostrara alguna particularidad.

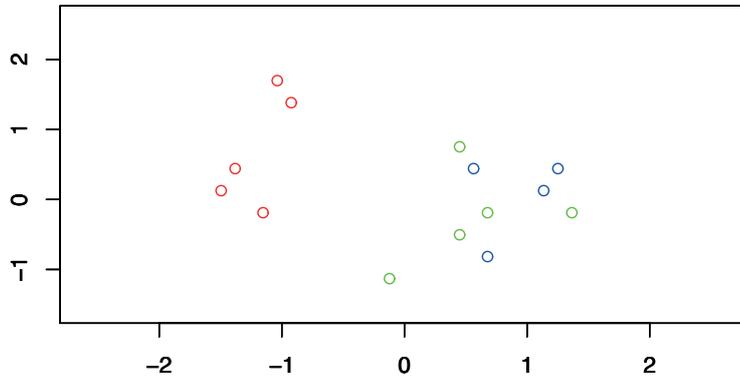
Como muestra de estos diagramas de dispersión, la figura 1 recoge las variables de longitud del sépalo y anchura del sépalo, para las quince flores. En el caso de la figura 2, se muestran las variables de anchura del sépalo y anchura del

Diagrama de dispersión

Un diagrama de dispersión (en inglés, *scatter plot*) muestra gráficamente la relación entre dos variables cuantitativas.

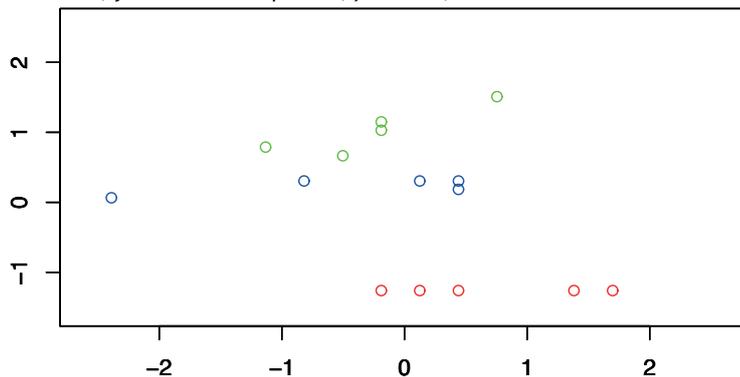
pétalo. Los colores representan el tipo o especie de flor: *setosa* (rojo), *versicolor* (azul) y *virginica* (verde). En las dos figuras se puede ver cómo los puntos azules se confunden con los puntos verdes, al mismo tiempo que los puntos rojos quedan agrupados de manera muy diferenciada.

Figura 1. Diagrama de dispersión de las variables relativas a la longitud del sépalo (eje horizontal) y la anchura del sépalo (eje vertical).



Los colores representan el tipo o la especie de flor: *setosa* (rojo), *versicolor* (azul) y *virginica* (verde).
Fuente: elaboración propia

Figura 2. Diagrama de dispersión de las variables relativas a la anchura del sépalo (eje horizontal) y la anchura del pétalo (eje vertical).



Los colores representan el tipo o la especie de flor: *setosa* (rojo), *versicolor* (azul) y *virginica* (verde).
Fuente: elaboración propia

2.2 Matriz de covarianzas

Dada la matriz (normalizada)

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} \in \mathcal{M}_{n \times m}(\mathbb{R})$$

$$= \begin{bmatrix} v_1 & v_2 & \cdots & v_j & \cdots & v_m \end{bmatrix}$$

la matriz de covarianzas se define así:

$$\mathbf{C}_X = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} = \frac{1}{n-1} \begin{bmatrix} v_1^T v_1 & v_1^T v_2 & \cdots & v_1^T v_j & \cdots & v_1^T v_m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_j^T v_1 & v_j^T v_2 & \cdots & v_j^T v_j & \cdots & v_j^T v_m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_m^T v_1 & v_m^T v_2 & \cdots & v_m^T v_j & \cdots & v_m^T v_m \end{bmatrix} \in \mathcal{M}_{m \times m}(\mathbb{R})$$

Fijaos en que la matriz de covarianzas es una matriz cuadrada de tantas filas y columnas como columnas tiene la matriz \mathbf{X} .

La matriz de covarianzas \mathbf{C}_X

$$\mathbf{C}_X = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} = \frac{1}{n-1} \begin{pmatrix} \boxed{v_1^T v_1} & v_1^T v_2 & \cdots & v_1^T v_j & \cdots & v_1^T v_m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_j^T v_1 & v_j^T v_2 & \cdots & \boxed{v_j^T v_j} & \cdots & v_j^T v_m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_m^T v_1 & v_m^T v_2 & \cdots & v_m^T v_j & \cdots & \boxed{v_m^T v_m} \end{pmatrix}$$

mide el grado de relación lineal del conjunto de datos entre cada uno de los pares de variables. Los términos de la diagonal principal corresponden a la varianza de cada una de las variables:

$$\sigma_j^2 = \frac{1}{n-1} v_j^T v_j = \frac{1}{n-1} \sum_{i=1}^n x_{ij}^2$$

Dado que nuestros datos están normalizados, todos los términos de la diagonal principal son 1. Los términos que no son de la diagonal principal representan la covarianza entre cada par de variables:

$$\sigma_{jk}^2 = \frac{1}{n-1} v_j^T v_k = \frac{1}{n-1} \sum_{i=1}^n x_{ij} x_{ik}$$

En nuestro ejemplo, la matriz de covarianzas es

$$C_{\mathbf{X}} = \begin{bmatrix} 1.0000000 & -0.1609042 & 0.8854496 & 0.8251793 \\ -0.1609042 & 1.0000000 & -0.3817905 & -0.3578668 \\ 0.8854496 & -0.3817905 & 1.0000000 & 0.9860398 \\ 0.8251793 & -0.3578668 & 0.9860398 & 1.0000000 \end{bmatrix}$$

Se puede observar lo siguiente:

- 1) Los elementos de la diagonal principal son todos igual a 1. Esto es así porque nuestros datos han sido normalizados y, por lo tanto, la varianza de todos es 1.
- 2) La matriz de covarianzas es una matriz simétrica. Esto sucede porque la covarianza es simétrica, es decir,

$$\sigma_{jk}^2 = \sigma_{kj}^2.$$

- 3) Las variables 3 y 4, correspondientes a la longitud y anchura del pétalo, están altamente relacionadas, puesto que su covarianza es 0.9860398 (muy cercana a 1). Las variables 1 y 3 —longitud del sépalo y del pétalo, respectivamente— también están significativamente relacionadas, a pesar de que en menor proporción, puesto que su covarianza es 0.8854496.
- 4) Por el contrario, las variables 1 y 2 —longitud y anchura del sépalo, respectivamente— no están muy relacionadas. En efecto, su covarianza es -0.1609042 .

2.3 Diagonalización de la matriz de covarianzas

El objetivo del análisis de componentes principales es encontrar una transformación (aplicación) lineal:

$$\mathbf{P} \in \mathcal{M}_{m \times m}(\mathbb{R})$$

tal que los datos originales recogidos en \mathbf{X} se transformen o se proyecten en un nuevo espacio mediante el producto:

$$\mathbf{T} = \mathbf{XP} \in \mathcal{M}_{n \times m}(\mathbb{R})$$

de forma que la matriz de covarianzas $C_{\mathbf{T}}$ de los nuevos datos \mathbf{T} sea diagonal.

Como C_X es una matriz cuadrada y simétrica de dimensión $m \times m$, sabemos por el módulo “Aplicaciones lineales, diagonalización y vectores propios” que existen m valores propios λ_i reales y m vectores propios (ortonormales) p_i que forman una base en el espacio vectorial euclídeo \mathbb{R}^m tal que

$$C_X = PDP^T,$$

donde

$$P = \left[\begin{array}{c|c|c|c} p_1 & p_2 & \cdots & p_m \end{array} \right]$$

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$$

Dada la matriz X que contiene los datos originales (normalizados), los datos de la nueva matriz transformada T se calculan así:

$$T = XP \in \mathcal{M}_{n \times m}(\mathbb{R}),$$

donde P es la matriz en la que las columnas son los vectores propios de la matriz de covarianzas C_X .

Vectores ortonormales

Decimos que dos vectores p_i y p_j son ortonormales si $p_i^T p_j = 0$ y, además, $p_i^T p_i = 1$ y $p_j^T p_j = 1$.

Sobre la matriz P

Se puede demostrar fácilmente que la transpuesta de la matriz P es, a la vez, su inversa. Es decir, $P^T = P^{-1}$.

2.4 La matriz de covarianzas de los datos transformados

¿Cuál es la matriz de covarianzas de los datos transformados? Calculemosla:

$$\begin{aligned} C_T &= \frac{1}{n-1} T^T T = \frac{1}{n-1} P^T X^T X P = P^T C_X P \\ &= P^T P D P^T P = D = \text{diag}(\lambda_1, \dots, \lambda_m) \end{aligned}$$

Esto significa que la matriz de covarianzas de los datos transformados es diagonal. Por lo tanto, las nuevas variables están incorrelacionadas. Es habitual ordenar los vectores propios en función del valor propio asociado, de mayor a menor. Es decir, si los valores propios son:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m,$$

ubicaremos los vectores propios en la matriz P en este orden:

$$p_1, p_2, \dots, p_m.$$

Volviendo al ejemplo del conjunto de datos *Iris*, los valores propios y los vectores propios de la matriz de covarianzas C_X son:

$$\lambda_1 = 2.941490992, \quad p_1^T = \begin{bmatrix} 0.5260194 & -0.2616562 & 0.5786532 & 0.5656856 \end{bmatrix}$$

$$\lambda_2 = 0.891699528 \quad p_2^T = \begin{bmatrix} -0.31992346 & -0.94247968 & -0.06660176 & -0.07032238 \end{bmatrix}$$

$$\lambda_3 = 0.162361649 \quad p_3^T = \begin{bmatrix} 0.7635779 & -0.2005818 & -0.2275748 & -0.5700223 \end{bmatrix}$$

$$\lambda_4 = 0.004447831 \quad p_4^T = \begin{bmatrix} 0.1946827 & -0.0550912 & -0.7803425 & 0.5917171 \end{bmatrix}$$

El vector propio asociado al valor propio más grande se denomina *primera componente principal*. El vector propio asociado al segundo valor propio más grande se denomina *segunda componente principal*. Y así sucesivamente.

2.4.1 ¿Cómo debemos interpretar, por ejemplo, la primera componente principal?

Denotamos las cuatro variables que hemos considerado en el ejemplo de las flores como u_1, u_2, u_3 y u_4 , donde

- 1) u_1 es la longitud del sépalo;
- 2) u_2 es la anchura del sépalo;
- 3) u_3 es la longitud del pétalo, y
- 4) u_4 es la anchura del pétalo.

Las componentes de la primera componente principal p_1 indican que

$$p_1 = 0.5260194u_1 - 0.2616562u_2 + 0.5786532u_3 + 0.5656856u_4.$$

En otras palabras, p_1 representa una nueva variable, que es una combinación lineal de las cuatro variables originales. Fijaos en que en la definición de esta nueva variable no todas las variables originales tienen el mismo peso. En efecto, la variable que tiene más peso es u_3 (longitud del pétalo), seguida de la variable u_4 (anchura del pétalo). Claramente, la que tiene menos influencia en la primera componente es la variable u_2 (anchura del sépalo).

¿Cuál es el valor de la nueva variable p_1 para el caso de la primera flor de la tabla 3? Recordad que las variables normalizadas en el caso de la primera flor son:

$$u_1 = -0.923772415$$

$$u_2 = 1.383518809$$

$$u_3 = -1.307336408$$

$$u_4 = -1.261519508$$

Por lo tanto, para la primera flor

$$p_1 = 0.5260194u_1 - 0.2616562u_2 + 0.5786532u_3 + 0.5656856u_4 = -2.3177306$$

Si hacemos lo mismo con las otras componentes principales, obtenemos

$$p_2 = -0.8328099$$

$$p_3 = 0.03418822$$

$$p_4 = 0.017762021$$

En el caso general, las medidas en las nuevas variables se obtienen de multiplicar los datos originales de la matriz \mathbf{X} por la matriz \mathbf{P} de las componentes principales:

$$\mathbf{T} = \mathbf{XP} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1j} & \cdots & t_{1m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \cdots & t_{ij} & \cdots & t_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nj} & \cdots & t_{nm} \end{bmatrix} \in \mathcal{M}_{n \times m}(\mathbb{R})$$

Fijaos en que las matrices \mathbf{X} y \mathbf{T} tienen la misma dimensión.

La matriz T para el ejemplo de las flores sería:

$$T = \begin{bmatrix} -2.31773055645763 & -0.832809881637843 & 0.0341882231504618 & 0.0177620212434633 \\ -2.02675817141067 & 0.722164171113012 & 0.174766954611649 & 0.0598152395437477 \\ -2.34229080476915 & 0.206211215432147 & -0.114116476558486 & 0.021856594229762 \\ -2.25904377241743 & 0.53213302544293 & -0.162490314930362 & -0.0655943343449898 \\ -2.46020386002701 & -1.09254655308811 & -0.116267982265906 & -0.0218406874991054 \\ 0.900381739958164 & -0.857289184578846 & 0.662337916991262 & -0.0122754574068039 \\ 0.546195294147313 & -0.639046440726892 & 0.0933763504900945 & 0.00785590242460743 \\ 1.05180934990985 & -0.539843144155038 & 0.545260806182149 & -0.0284082915673973 \\ 0.455578827059034 & 2.39217029459659 & 0.0720848822574889 & 0.0268371163600159 \\ 0.966077049158865 & 0.506218727354301 & 0.4210140657984 & 0.0581798257612053 \\ 1.54431815987357 & -1.03634661634201 & -0.924547274237649 & 0.0627325552563454 \\ 1.05260949959829 & 1.00735403635774 & -0.462576054699356 & -0.00141932059541922 \\ 1.96942554310489 & -0.402777918923283 & 0.250593622580613 & 0.0489153084361122 \\ 1.27380657875815 & 0.222463831446303 & -0.143227795787201 & -0.202216367937914 \\ 1.64582512351377 & -0.188055562291003 & -0.330396923583155 & 0.0277998960963719 \end{bmatrix}$$

2.4.2 ¿Cuál es el peso de la primera componente principal?

Por las características de la matriz de covarianzas C_X —simétrica y definida positiva—, todos los valores propios son positivos. Además, podemos observar que en el ejemplo de las flores

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 4,$$

es decir,

$$\sum_{i=1}^m \lambda_i = \text{tr}(C_X)$$

Ya hemos visto que los elementos de la diagonal principal de la matriz de covarianzas C_X representan el valor de la varianza de cada una de las variables originales. En el caso de las nuevas variables, los elementos de la diagonal principal de la matriz de covarianzas C_T —que son los valores propios λ_i — también representan esta varianza. Si el total de la varianza, es decir, la suma de los valores propios, es m , la aportación de la primera componente es

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \times 100 \%$$

Traza de una matriz

Recordad que la traza de una matriz cuadrada, que denotamos por $\text{tr}(A)$, es la suma de los elementos de la diagonal principal.

Para nuestro ejemplo, la primera componente principal es capaz de retener un porcentaje de variabilidad igual a:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \times 100 \% = \frac{2.941490992}{4} \times 100 \% = 73.53727480 \%$$

Del mismo modo, el resto de las componentes principales son capaces de retener el porcentaje de variabilidad siguiente:

$$\frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \times 100 \% = \frac{0.891699528}{4} \times 100 \% = 22.29248820 \%$$

$$\frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \times 100 \% = \frac{0.162361649}{4} \times 100 \% = 4.059041225 \%$$

$$\frac{\lambda_4}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \times 100 \% = \frac{0.004447831}{4} \times 100 \% = 0.1111957750 \%$$

Cada una de las cuatro variables originales u_1, u_2, u_3 y u_4 retiene un 25 % de la variabilidad. En cambio, las nuevas variables p_1, p_2, p_3 y p_4 —las cuatro componentes principales— retienen un 73.5 %, 22.3 %, 4.1 % y 0.1 %, respectivamente. Es decir, hacen falta tres variables originales para obtener la misma cantidad de información que se conseguiría con una única variable nueva, la primera componente principal.

Variabilidad

Debemos entender la variabilidad como la cantidad de información. Cuanta más variabilidad hay, más información tenemos.

Tabla 4. Variabilidad acumulada por las cuatro variables originales y las cuatro variables nuevas (componentes principales).

variables	variabilidad acumulada (variables originales)	variabilidad acumulada (componentes principales)
1	25 %	73.5 %
2	50 %	95.8 %
3	75 %	99.9 %
4	100 %	100 %

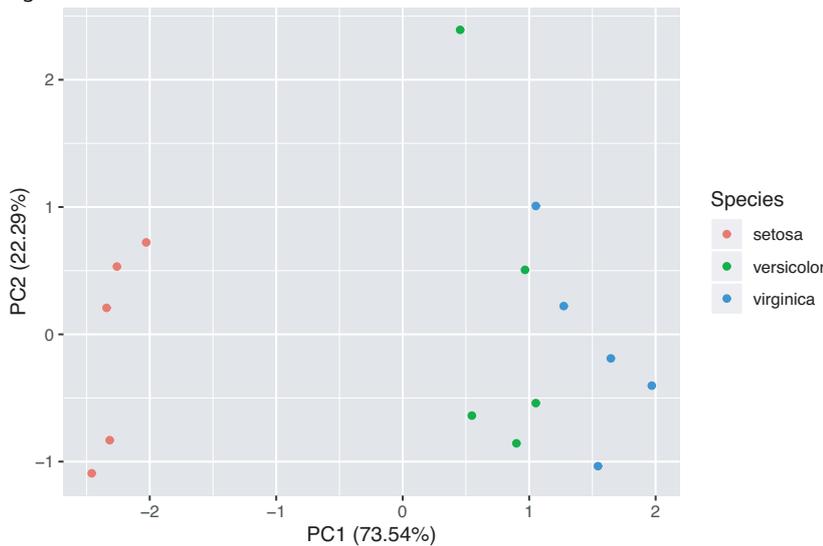
Fuente: elaboración propia

Observemos ahora las figuras 3 y 4, que contienen información interesante:

- 1) Las dos figuras contienen la proyección de las quince flores sobre las dos primeras componentes principales. En este caso, el código de colores es: *setosa* (rojo), *versicolor* (verde) y *virginica* (azul). El grupo *setosa* (rojo) continúa claramente diferenciado. Al mismo tiempo, la separación entre el grupo de flores *versicolor* (verde) y *virginica* (azul) ahora es más clara.
- 2) La figura 4 contiene, además, la contribución de cada una de las cuatro variables originales a las dos primeras componentes principales. Si miramos, por ejemplo, la primera componente principal (la dirección horizontal), las variables que más intervienen son la longitud del sépalo (*Sepal.Length*) y la longitud y anchura del pétalo (*Petal.Length*, *Petal.Width*). En el caso de la segunda componente principal (la dirección vertical), la variable que tiene más peso es la anchura del sépalo (*Sepal.Width*).

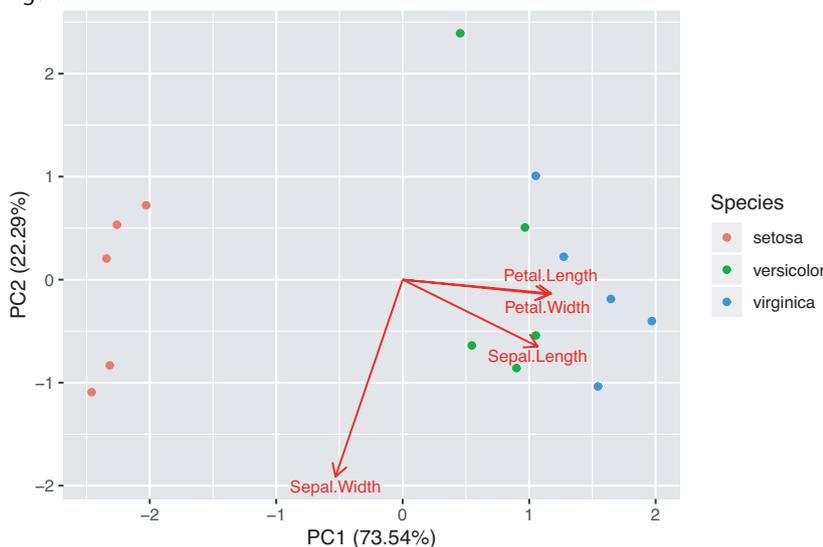
3) En la figura 4 también se puede ver cómo las flechas que indican las direcciones de las variables longitud y anchura del pétalo (*Petal.Length*, *Petal.Width*) están prácticamente superpuestas. Recordemos que, en este caso, la covarianza entre estas dos variables es 0.9860398, lo que representa un valor muy cercano a 1. Es decir, como ya habíamos dicho las variables 3 y 4 están altamente relacionadas.

Figura 3



Fuente: elaboración propia

Figura 4



Fuente: elaboración propia

2.5 Reducción de la dimensión

En la tabla 4 hemos visto que, en el ejemplo de las flores, con dos componentes principales podemos retener el 95.8 % de la variabilidad o la información. Esto significa que podemos pasar de una muestra de quince flores en que he-

mos medido cuatro variables diferentes a una muestra de quince flores en que solo medimos dos variables. Veámoslo con más detalle en el caso general.

Si consideramos todas las componentes principales, la ortonormalidad de las componentes principales implica que

$$\mathbf{P}\mathbf{P}^T = \mathbf{I}_m,$$

donde \mathbf{I}_m es la matriz identidad de dimensión m . Entonces, la proyección:

$$\mathbf{T} = \mathbf{X}\mathbf{P}$$

se puede invertir para recuperar los datos originales, a partir de los datos proyectados:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T,$$

puesto que

$$\mathbf{T} = \mathbf{X}\mathbf{P} \Leftrightarrow \mathbf{T}\mathbf{P}^T = \mathbf{X}\mathbf{P}\mathbf{P}^T \Leftrightarrow \mathbf{T}\mathbf{P}^T = \mathbf{X}\mathbf{I}_m \Leftrightarrow \mathbf{T}\mathbf{P}^T = \mathbf{X}.$$

Sin embargo, uno de los objetivos del análisis de componentes principales es la reducción de la dimensión. Por eso, consideramos ahora un número inferior de componentes principales, $\ell < m$, es decir, solo consideramos los vectores propios asociados a los ℓ valores propios más grandes. Entonces, si definimos la matriz reducida:

$$\hat{\mathbf{P}} = (p_1 | p_2 | \dots | p_\ell) \in \mathcal{M}_{m \times \ell}(\mathbb{R})$$

la matriz de las proyecciones se define así:

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{P}} \in \mathcal{M}_{n \times \ell}(\mathbb{R})$$

2.5.1 El error residual

Una de las consecuencias de haber reducido la dimensionalidad es que la matriz $\hat{\mathbf{P}}$ ya no es invertible. Por lo tanto, los datos originales contenidos en X no se pueden recuperar completamente mediante la matriz $\hat{\mathbf{T}}$. Sin embargo, se puede invertir la proyección de la manera siguiente:

$$\hat{X} = \hat{\mathbf{T}}\hat{\mathbf{P}}^T \in \mathcal{M}_{n \times m}(\mathbb{R})$$

para obtener los datos originales con pérdida de información. La diferencia entre los datos originales recogidos en la matriz X y los datos originales con pérdida de información de la matriz \hat{X} se denomina *error residual* y se representa con la matriz E . En efecto:

$$E = X - \hat{X} \in \mathcal{M}_{n \times m}(\mathbb{R})$$

Sobre la dimensión de $\hat{\mathbf{T}}$

$\hat{\mathbf{T}}$ es una matriz que continúa teniendo tantas filas como la matriz \mathbf{X} original. De todas formas, el número de columnas de la matriz $\hat{\mathbf{T}}$ pasa de tener m columnas a tener ℓ columnas.

Nota

La matriz $\hat{\mathbf{P}}$ ya no es invertible. De hecho, la matriz $\hat{\mathbf{P}}$ no es, ni siquiera, cuadrada.

2.6 Ejemplo de aplicación: encuesta de presupuestos familiares

Continuamos con el ejemplo del subapartado 1.2, en el que se presentan los datos de la encuesta de presupuestos familiares (EPF) por comunidades autónomas y las variables que se miden. Como no hay información de la variable *Enseñanza* para Ceuta y Melilla, eliminamos estas dos ciudades autónomas del análisis. Queremos ver si, gracias al análisis de componentes principales, podemos extraer alguna conclusión o destacar algún patrón que haya quedado oculto en la cantidad de datos.

Con la ayuda del lenguaje de programación R, los datos han sido almacenados en la matriz `INE`, que contiene diecisiete filas —una por comunidad autónoma— y doce columnas —una por cada variable. Procedemos de la manera siguiente, tal como se puede ver en la figura 5:

1) Con la instrucción `prcomp`, calculamos el análisis de componentes principales, que almacenamos en la variable `ine.pca`. Es importante añadir las opciones `center = TRUE` y `scale = TRUE` para garantizar que los datos se han escalado y centrado.

2) La primera componente principal se obtiene con `ine.pca$rotation[,1]`. Se puede observar que la primera componente es una combinación lineal de las doce variables originales. En particular:

$$p_1 = 0.30u_1 + 0.12u_2 + 0.28u_3 + 0.33u_4 + 0.33u_5 + 0.31u_6 + 0.18u_7 + 0.25u_8 + 0.31u_9 + 0.30u_{10} + 0.33u_{11} + 0.33u_{12}$$

Figura 5. Instrucciones de R para el cálculo de las dos primeras componentes principales y los valores de las variables originales proyectados sobre estas dos primeras componentes.

```
> ine.pca <- prcomp(INE[,2:13],center = TRUE, scale=TRUE)
> ine.pca$rotation[,1]
      V1      V2      V3      V4      V5      V6      V7
0.3048175 0.1177153 0.2797855 0.3336785 0.3299060 0.3053362 0.1788030
      V8      V9     V10     V11     V12
0.2512488 0.3112978 0.3010705 0.3349005 0.3267037
> ine.pca$rotation[,2]
      V1      V2      V3      V4      V5
-0.01383321 -0.71634090 -0.36251483 0.18264741 -0.13857533
      V6      V7      V8      V9     V10
-0.07815930 0.08018175 0.44706142 -0.12810849 0.22574457
      V11     V12
-0.01588222 0.15052448
> ine.pca$x[,1]
 [1] -1.5284356 0.5230432 0.1936867 1.8834382 -4.5577019 1.4096488
 [7] -0.6333325 -4.0799862 2.4422616 -0.8963849 -4.9655611 -0.8830605
[13] 3.7092028 -0.8849228 3.6507548 4.3414574 0.2758919
> ine.pca$x[,2]
 [1] -1.38763977 -0.42984445 -0.20352852 -0.01118605 2.85476875
 [6] -0.01202456 -0.13105989 -0.46300999 1.14629871 -0.07641837
[11] 0.41923006 -1.32859163 0.47571073 -1.71830933 -0.45566480
[16] 1.29743154 0.02383759
```

Fuente: elaboración propia

Datos completos

Los datos completos de la encuesta de presupuestos familiares 2017 se pueden obtener en este enlace del Instituto Nacional de Estadística:
<https://www.ine.es/jaxiT3/Tabla.htm?t=25143&L=0>.

R

R es un entorno de programación libre especializado en estadística y representación gráfica.

La primera componente principal representa una nueva variable donde todas las variables originales *suman* en más o menos proporción. En particular, las variables que tienen más peso son: la u_4 (vivienda, agua, electricidad, gas y otros combustibles), la u_5 (muebles, artículos del hogar y artículos para el mantenimiento del hogar), la u_{11} (restaurantes y hoteles) y la u_{12} (otros bienes y servicios). Es especialmente relevante el poco peso que tienen las variables u_2 (bebidas alcohólicas y tabaco) y u_7 (transporte). La variable u_2 será, en cambio, la más importante en la segunda componente principal, como se verá a continuación. La variable u_7 continuará teniendo poca influencia. Como se puede ver en la figura 6, la información o variabilidad explicada por la primera componente principal es del 65.1%.

3) Si proyectamos los datos originales de las diecisiete comunidades autónomas sobre la primera componente principal, lo que podemos hacer con la orden `ine.pca$x[,1]`, obtendremos los resultados que se pueden ver en la figura 5. Si ordenamos estos valores de grandes a pequeños, obtendremos una primera ordenación interesante, que se puede ver en la tabla 5. Podríamos decir que la primera componente principal ha sido capaz de ordenar las comunidades autónomas en función de la renta.

4) La segunda componente principal se obtiene con `ine.pca$rotation[,2]`. Se puede observar que la segunda componente es también una combinación lineal de las doce variables originales. En particular:

$$p_2 = -0.01u_1 - 0.72u_2 - 0.36u_3 + 0.18u_4 - 0.14u_5 - 0.08u_6 \\ + 0.08u_7 + 0.45u_8 - 0.13u_9 + 0.23u_{10} - 0.02u_{11} + 0.15u_{12}$$

Tabla 5. Proyección de los datos originales sobre la primera componente principal (de grande a pequeña)

Comunidad autónoma	PC1
País Vasco	4.3414574
Comunidad de Madrid	3.7092028
Navarra	3.6507548
Cataluña	2.4422616
Baleares	1.8834382
Cantabria	1.4096488
Aragón	0.5230432
La Rioja	0.2758919
Asturias	0.1936867
Castilla y León	-0.6333325
Galicia	-0.8830605
Murcia	-0.8849228
Comunidad Valenciana	-0.8963849
Andalucía	-1.5284356
Castilla-La Mancha	-4.0799862
Canarias	-4.5577019
Extremadura	-4.9655611

Fuente: Encuesta de presupuestos familiares 2017 (Instituto Nacional de Estadística)

En este caso, algunas de las variables originales suman, mientras que otras restan, en su contribución a la nueva variable que representa la segunda com-

ponente principal. Las variables con más peso (en valor absoluto) son: u_2 (bebidas alcohólicas y tabaco), u_8 (comunicaciones) y u_3 (vestido y calzado). También en esta segunda componente principal, el peso de la variable u_7 (transporte) es muy pequeño. Esto significa que la variable que mide los gastos en transporte no marca una diferencia entre las comunidades autónomas (podríamos obtener un resultado diferente si estudiáramos la encuesta de presupuestos familiares por provincias y no por comunidades autónomas). Como se puede ver en la figura 6, la información o variabilidad explicada por la segunda componente principal es del 9.83 %. Por lo tanto, la variabilidad explicada por las dos primeras componentes principales es del 74.93 %.

5) Si proyectamos los datos originales de las diecisiete comunidades autónomas sobre la segunda componente principal, lo que podemos hacer con la orden `ine.pca$x[,2]`, conseguiremos los resultados que se pueden ver en la figura 5. Si ordenamos de grandes a pequeños estos valores, obtendremos una segunda ordenación interesante, que se puede ver en la tabla 6. Sin embargo, en este caso es más difícil explicar de manera cualitativa cuál es la interpretación de esta segunda variable.

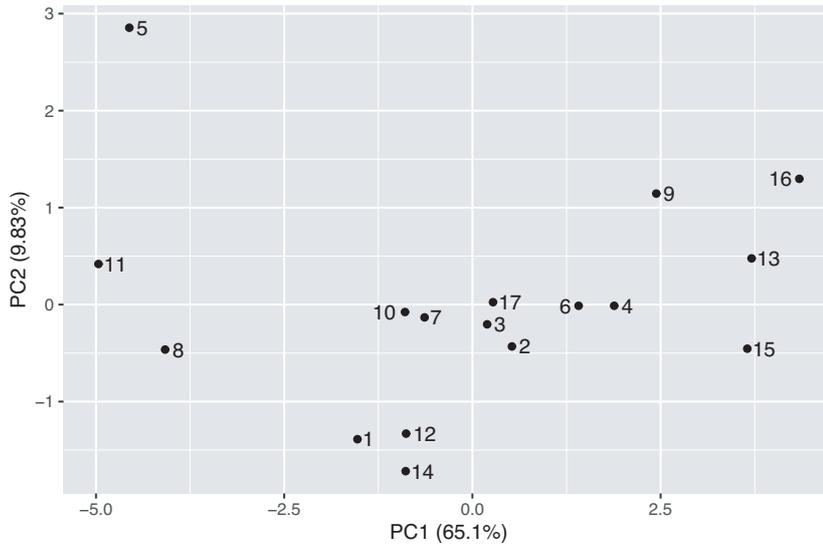
6) En la figura 6 se puede ver la proyección de los datos originales de las diecisiete comunidades autónomas sobre las dos primeras componentes principales, que son capaces de retener casi el 75 % de la información.

Tabla 6. Proyección de los datos originales sobre la segunda componente principal (de grande a pequeña)

Comunidad autónoma	PC2
Canarias	2.85476875
País Vasco	1.29743154
Cataluña	1.14629871
Comunidad de Madrid	0.47571073
Extremadura	0.41923006
La Rioja	0.02383759
Baleares	-0.01118605
Cantabria	-0.01202456
Comunidad Valenciana	-0.07641837
Castilla y León	-0.13105989
Asturias	-0.20352852
Aragón	-0.42984445
Navarra	-0.45566480
Castilla-La Mancha	-0.46300999
Galicia	-1.32859163
Andalucía	-1.38763977
Murcia	-1.71830933

Fuente: Encuesta de presupuestos familiares 2017 (Instituto Nacional de Estadística)

Figura 6. Proyección de los datos originales de las diecisiete comunidades autónomas sobre las dos primeras componentes principales, que son capaces de retener casi el 75 % de la información.



Fuente: elaboración propia

3. Descomposición en valores singulares

3.1 Ejemplo introductorio

Las matrices, como elemento básico del álgebra lineal, pueden ser utilizadas en un amplio espectro de aplicaciones, desde las más sencillas (sistema de ecuaciones lineales) hasta las más complejas (sistemas dinámicos discretos, transición de estados). Una de estas aplicaciones es emplear una matriz para almacenar los píxeles de una imagen. Cada elemento de la matriz representa un píxel. Si la imagen es en escala de grises, los valores que puede adoptar cada píxel estarían en el rango $[0,1]$.

Píxel

Según el *Diccionario de la Real Academia Española*, es la superficie homogénea más pequeña de las que componen una imagen, que se define por su brillo y color.

Consideramos, por ejemplo, una imagen de píxeles 6×6 almacenada en forma de matriz:

$$A = \begin{bmatrix} 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \end{bmatrix}$$

A partir de esta matriz podemos hacer las observaciones siguientes:

- 1) La matriz A tiene rango 1, puesto que todas las filas son iguales.
- 2) La matriz A está formada por 36 elementos, que es el resultado de multiplicar el número de filas (6) por el número de columnas (también 6).

Se puede ver fácilmente que la matriz A se puede expresar como el producto de un vector columna (u) por un vector fila (v^T) si consideramos que:

$$A = uv^T = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0.1 & 0.1 & 0.5 & 0.5 & 0.9 & 0.9 \end{bmatrix}$$

¿Cuál es la importancia de la descomposición de la matriz \mathbf{A} actual como producto de estos dos vectores? De entrada, que para transmitir la información contenida en estos dos vectores solo necesitamos:

$$6 \times 2 = 12$$

números, en lugar de los 36 necesarios para transmitir toda la matriz.

3.2 Descomposición en valores singulares reducida

A partir de una matriz $\mathbf{A} \in \mathcal{M}_{m \times n}(\mathbb{R})$:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

$\mathcal{M}_{m \times n}(\mathbb{R})$

Recordad que denotamos por medio de $\mathcal{M}_{m \times n}(\mathbb{R})$ el espacio vectorial de las matrices de m filas y n columnas a coeficientes reales.

la descomposición en valores singulares (SVD) **reducida** de \mathbf{A} es una factorización (o descomposición)

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

donde

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{n1} \\ u_{12} & u_{22} & \cdots & u_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1m} & u_{2m} & \cdots & u_{nm} \end{bmatrix} \in \mathcal{M}_{m \times n}(\mathbb{R}) \text{ es una matriz ortogonal,}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} & \cdots & v_{n1} \\ v_{12} & v_{22} & \cdots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \cdots & v_{nn} \end{bmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R}) \text{ es una matriz ortogonal y}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R}) \text{ es una matriz diagonal.}$$

Los valores $\sigma_i \geq 0$, $i = 1, \dots, n$ se denominan **valores singulares** y en general se enumeran de manera descendente:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

Los vectores \mathbf{u}_i , $i = 1, \dots, n$ se denominan **vectores singulares por la izquierda** y, de manera similar, los vectores \mathbf{v}_i , $i = 1, \dots, n$ se denominan **vectores singulares por la derecha**.

El hecho de que tanto la matriz \mathbf{U} como la matriz \mathbf{V} sean ortogonales significa que

$$\mathbf{u}_i^T \mathbf{u}_j = 0, \quad i \neq j,$$

$$\mathbf{u}_i^T \mathbf{u}_i = 1,$$

$$\mathbf{v}_i^T \mathbf{v}_j = 0, \quad i \neq j,$$

$$\mathbf{v}_i^T \mathbf{v}_i = 1.$$

3.2.1 Cálculo de los valores singulares y de los vectores singulares

Supongamos que la matriz \mathbf{A} tiene rango r . Para calcular los valores singulares y los vectores singulares, haremos lo siguiente:

- 1) Los valores singulares no nulos σ_i , con $i \leq r$, son la raíz cuadrada de los valores propios no nulos de las matrices $\mathbf{A}^T \mathbf{A}$ y $\mathbf{A} \mathbf{A}^T$.
- 2) Los vectores singulares por la derecha de la matriz \mathbf{V} son los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$ (base ortonormal).
- 3) Los vectores singulares por la izquierda de la matriz \mathbf{U} son los vectores propios de la matriz $\mathbf{A} \mathbf{A}^T$ (base ortonormal).
- 4) Usando estas dos bases, \mathbf{A} se convierte en una matriz diagonal $\mathbf{\Sigma}$ y se obtiene:

$$\mathbf{A} \mathbf{v}_i = \sigma_i \mathbf{u}_i,$$

donde σ_i son los valores singulares.

Los vectores singulares son únicos, salvo signo. Por lo tanto, se propone la secuencia siguiente para encontrar los vectores propios por la derecha y por la izquierda:

- 1) Cálculo de los valores singulares, que son las raíces cuadradas de los valores propios de la matriz $\mathbf{A}^T \mathbf{A}$ —y $\mathbf{A} \mathbf{A}^T$.
- 2) Cálculo de los vectores singulares por la derecha, \mathbf{v}_i , que son los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$.
- 3) Cálculo de los vectores singulares por la izquierda, \mathbf{u}_i , teniendo en cuenta la relación

$$\mathbf{u}_i = \mathbf{A} \mathbf{v}_i / \sigma_i,$$

de forma que aseguramos que los signos están escogidos correctamente.

Ejemplo

Considerad la matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}.$$

Las matrices $\mathbf{A}^T \mathbf{A}$ y $\mathbf{A} \mathbf{A}^T$ son:

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\mathbf{A} \mathbf{A}^T = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$$

Las dos matrices tienen los mismos valores propios. Como $\mathbf{A} \mathbf{A}^T$ ya es una matriz diagonal, los valores propios son $\lambda_1 = 8$ y $\lambda_2 = 2$. Por lo tanto, los valores singulares son:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{8} = 2\sqrt{2}$$

$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$$

Para encontrar los vectores propios de la matriz $\mathbf{A} \mathbf{A}^T$, teniendo en cuenta que los valores propios son $\lambda_1 = 8$ y $\lambda_2 = 2$, hay que resolver estas ecuaciones:

$$\mathbf{A} \mathbf{A}^T \mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

$$\mathbf{A} \mathbf{A}^T \mathbf{u}_2 = \lambda_2 \mathbf{u}_2$$

En el caso de la primera, tenemos lo siguiente:

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_1 = \lambda_1\mathbf{u}_1 \Leftrightarrow \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = 8 \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} \Leftrightarrow \begin{cases} u_{11} = \kappa, \\ -6u_{12} = 0 \end{cases}$$

El sistema anterior es un sistema compatible indeterminado, donde κ es un parámetro y $u_{12} = 0$. Dado que queremos imponer que el vector \mathbf{u}_1 tenga norma 1, proponemos, por ejemplo, $u_{11} = 1$. Por lo tanto:

$$\mathbf{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Procederemos de una manera similar en el caso de la segunda ecuación asociada al valor propio $\lambda_2 = 2$; el vector propio \mathbf{u}_2 que obtenemos es:

$$\mathbf{u}_2 = \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Para encontrar los vectores propios de la matriz $\mathbf{A}^T\mathbf{A}$, teniendo en cuenta que los valores propios son $\lambda_1 = 8$ y $\lambda_2 = 2$, hay que resolver estas ecuaciones:

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$$

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_2 = \lambda_2\mathbf{v}_2$$

En el caso de la primera, tenemos lo siguiente:

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \Leftrightarrow \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 8 \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \Leftrightarrow \begin{cases} -3v_{11} + 3v_{12} = 0, \\ 3v_{11} - 3v_{12} = 0 \end{cases}$$

El sistema anterior es un sistema compatible indeterminado, con solución paramétrica $v_{11} = v_{12}$. Dado que queremos imponer que el vector \mathbf{v}_1 tenga norma 1, proponemos, por ejemplo, $v_{11} = v_{12} = 1/\sqrt{2}$. Por lo tanto:

$$\mathbf{v}_1 = \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Procederemos de una manera similar en el caso de la segunda ecuación asociada al valor propio $\lambda_2 = 2$; el vector propio \mathbf{v}_2 que obtenemos es:

$$\mathbf{v}_2 = \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Finalmente, fijaos en que:

$$\begin{aligned} \mathbf{A} &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T \\ &= \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \end{aligned}$$

Se puede observar, además, que, a partir de una matriz $\mathbf{A} \in \mathcal{M}_{m \times n}(\mathbb{R})$ de rango r :

- 1) Los vectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ forman una base ortonormal del subespacio generado por las **filas** de la matriz \mathbf{A} .
- 2) Los vectores $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n$ forman una base ortonormal del subespacio generado por el **núcleo** de la matriz \mathbf{A} .
- 3) Los vectores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ forman una base ortonormal del subespacio generado por **columnas** de la matriz \mathbf{A} .
- 4) Los vectores $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_m$ forman una base ortonormal del subespacio generado por el **núcleo** de la matriz \mathbf{A}^T .

El comentario anterior se entenderá mejor si consideramos el ejemplo siguiente:

Ejemplo

Queremos calcular la descomposición en valores singulares de la matriz:

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}.$$

La primera observación que podemos hacer es que el rango de la matriz es $r = 1$, puesto que el determinante de la matriz es cero. Por lo tanto, la dimensión del subespacio generado por las filas de la matriz \mathbf{A} es igual que $r = 1$. Así pues, una base de este subespacio estará formada por un único vector, que será el primer vector singular por la derecha \mathbf{v}_1^T y será igual que el vector de la primera fila de la matriz, si lo normalizamos. En efecto:

$$\mathbf{v}_1^T = [4, 3] / \sqrt{4^2 + 3^2} = [4, 3] / 5 = [4/5, 3/5]$$

Dado que el número de columnas es $n = 2$, la dimensión del subespacio generado por el núcleo es $n - r = 2 - 1 = 1$. Así pues, cualquier vector ortogonal a \mathbf{v}_1 será la base de este subespacio. Así pues, simplemente hay que considerar

$$\mathbf{v}_2^T = [-u_{12}, u_{11}] = [-3/5, 4/5],$$

Ejemplo ilustrativo

Este ejemplo ilustra una manera de calcular los vectores singulares por la izquierda y por la derecha sin tener que calcular vectores propios.

puesto que

$$\mathbf{v}_1^T \mathbf{v}_2 = 0.$$

Además, sabemos que

$$\mathbf{A}\mathbf{v}_1 = \sigma_1 \mathbf{u}_1.$$

Por lo tanto:

$$\mathbf{A}\mathbf{v}_1 = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix} = \sqrt{125} \begin{bmatrix} 5/\sqrt{125} \\ 10/\sqrt{125} \end{bmatrix} = 5\sqrt{5} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Así pues, $\sigma_1 = 5\sqrt{5}$ y $\mathbf{u}_1^T = [1/\sqrt{5}, 2/\sqrt{5}]$. Adicionalmente,

$$\mathbf{u}_2^T = [-2/\sqrt{5}, 1/\sqrt{5}].$$

En este caso, también resulta fácil comprobar que

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T$$

Es fácil comprobar que \mathbf{U} y \mathbf{V} son los vectores propios de valor propio σ_1^2 y $\sigma_2^2 = 0$ de las matrices $\mathbf{A}\mathbf{A}^T$ y $\mathbf{A}^T\mathbf{A}$, respectivamente.

3.3 Descomposición en valores singulares completa

A partir de una matriz $\mathbf{A} \in \mathcal{M}_{m \times n}(\mathbb{R})$:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

SVD completa

La diferencia entre la SVD reducida y la SVD completa es que la base ortonormal \mathbf{U} formada por n elementos en el caso de la SVD reducida se amplía a una base ortonormal completa de \mathbb{R}^n . Con ello, la matriz Σ se amplía a una matriz rectangular $m \times n$ añadiendo ceros.

la descomposición en valores singulares (SVD) **completa** de \mathbf{A} es una factorización (o descomposición)

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T,$$

donde

$$\begin{aligned}
 \mathbf{U} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \\
 &= \left[\begin{array}{cccc|cccc} u_{11} & u_{21} & \cdots & u_{n1} & u_{n+1,1} & \cdots & u_{m1} & \\ u_{12} & u_{22} & \cdots & u_{n2} & u_{n+1,2} & \cdots & u_{m2} & \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ u_{1m} & u_{2m} & \cdots & u_{nm} & u_{n+1,m} & \cdots & u_{mm} & \end{array} \right] \in \mathcal{M}_{m \times m}(\mathbb{R}) \text{ es una matriz ortogonal cuadrada de orden } m,
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{V} &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \\
 &= \left[\begin{array}{cccc} v_{11} & v_{21} & \cdots & v_{n1} \\ v_{12} & v_{22} & \cdots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \cdots & v_{nn} \end{array} \right] \in \mathcal{M}_{n \times n}(\mathbb{R}) \text{ es una matriz ortogonal cuadrada de orden } n \text{ y}
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{\Sigma} &= \left[\begin{array}{cccc|cccc} \sigma_1 & 0 & \cdots & 0 & & & & \\ 0 & \sigma_2 & \cdots & 0 & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & \\ 0 & 0 & \cdots & \sigma_n & & & & \\ \hline 0 & 0 & \cdots & 0 & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & \\ 0 & 0 & \cdots & 0 & & & & \end{array} \right] \in \mathcal{M}_{m \times n}(\mathbb{R}) \text{ es una matriz diagonal.}
 \end{aligned}$$

Los valores $\sigma_i \geq 0$, $i = 1, \dots, n$ se denominan **valores singulares** y en general se enumeran de manera descendente:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Los vectores \mathbf{u}_i , $i = 1, \dots, m$, como antes, se denominan **vectores singulares por la izquierda** y, de manera similar, los vectores \mathbf{v}_i , $i = 1, \dots, n$ se denominan **vectores singulares por la derecha**.

Consideraremos ahora un ejemplo en el que la matriz \mathbf{A} ya no será cuadrada, como en los dos ejemplos anteriores.

Ejemplo

Queremos calcular la descomposición en valores singulares (SVD) completa de la matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix}$$

Fijaos en que el número de filas de la matriz es $m = 3$ y que el número de columnas es $n = 2$. El rango máximo de la matriz es, por lo tanto, 2, puesto que es el mínimo entre estas dos cantidades:

$$\min(m,n) = \min(3,2) = 2.$$

El rango de la matriz es $r = 2$, puesto que la submatriz cuadrada 2×2

$$\begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$

tiene determinante no nulo. Para el cálculo de los vectores singulares por la derecha consideramos la matriz $\mathbf{A}^T \mathbf{A}$. En efecto:

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 8 \\ 8 & 9 \end{bmatrix}$$

Los valores propios de la matriz $\mathbf{A}^T \mathbf{A}$ son $\lambda_1 = 17$ y $\lambda_2 = 1$, de forma que los valores singulares son:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{17}$$

$$\sigma_2 = \sqrt{\lambda_2} = 1$$

Los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$ tienen que satisfacer:

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_1 = 17 \mathbf{v}_1$$

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_2 = \mathbf{v}_2$$

Si resolvemos las ecuaciones correspondientes, obtendremos:

$$\mathbf{v}_1^T = [1/\sqrt{2}, 1/\sqrt{2}]$$

$$\mathbf{v}_2^T = [1/\sqrt{2}, -1/\sqrt{2}]$$

Podemos encontrar los vectores singulares por la derecha teniendo en cuenta que:

$$\mathbf{A} \mathbf{v}_1 = \sigma_1 \mathbf{u}_1 = \sqrt{17} \mathbf{u}_1$$

$$\mathbf{A} \mathbf{v}_2 = \sigma_2 \mathbf{u}_2 = \mathbf{u}_2$$

En el primer caso,

$$\mathbf{A}\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{2} \\ 4/\sqrt{2} \\ 3/\sqrt{2} \end{bmatrix} = \sqrt{17} \begin{bmatrix} 3/\sqrt{34} \\ 4/\sqrt{34} \\ 3/\sqrt{34} \end{bmatrix}$$

y obtenemos

$$\mathbf{u}_1^T = [3/\sqrt{34}, 4/\sqrt{34}, 3/\sqrt{34}]$$

El vector \mathbf{u}_2 lo obtenemos de manera similar y el resultado es:

$$\mathbf{u}_2^T = [-1/\sqrt{2}, 0, 1/\sqrt{2}]$$

Adicionalmente, para la descomposición en valores singulares completa, necesitamos un tercer vector,

$$\mathbf{u}_3^T = [u_{31}, u_{32}, u_{33}],$$

que sea ortogonal en \mathbf{u}_1 y \mathbf{u}_2 . Entonces, el resultado es un sistema de ecuaciones compatible indeterminado:

$$\begin{aligned} 3u_{31} + 4u_{32} + 3u_{33} &= 0 \\ -u_{31} + u_{33} &= 0 \end{aligned}$$

que tiene solución paramétrica

$$\begin{aligned} u_{31} &= \kappa \\ u_{32} &= -3\kappa/2 \\ u_{33} &= \kappa \end{aligned}$$

Por ejemplo, podemos considerar

$$\mathbf{u}_3^T = [2/\sqrt{17}, -3/\sqrt{17}, 2/\sqrt{17}].$$

A continuación podemos mostrar tanto la descomposición en valores singulares completa como la reducida de la matriz \mathbf{A} :

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

$$\begin{aligned} &= \begin{bmatrix} 3/\sqrt{34} & -1/\sqrt{2} \\ 4/\sqrt{34} & 0 \\ 3/\sqrt{34} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{17} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T \\ &= \begin{bmatrix} 3/\sqrt{34} & -1/\sqrt{2} & 2/\sqrt{17} \\ 4/\sqrt{34} & 0 & -3/\sqrt{17} \\ 3/\sqrt{34} & 1/\sqrt{2} & 2/\sqrt{17} \end{bmatrix} \begin{bmatrix} \sqrt{17} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T \end{aligned}$$

3.3.1 Propiedades interesantes de la descomposición en valores singulares

- 1) Los valores singulares σ_i son únicos.
- 2) El rango de la matriz \mathbf{A} es el número de valores singulares no nulos.
- 3) $\|\mathbf{A}\|_2 = \sigma_1$ y $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$, donde

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}$$

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

- 4) Si \mathbf{A} es una matriz simétrica $\mathbf{A} = \mathbf{A}^T$, entonces los valores singulares de la matriz \mathbf{A} son el valor absoluto de los valores propios de la matriz \mathbf{A} . En particular, si \mathbf{A} es una matriz simétrica y definida positiva, los valores propios y los valores singulares coinciden.
- 5) Si \mathbf{A} es una matriz cuadrada, entonces el valor absoluto del determinante de la matriz coincide con el producto de los valores singulares, es decir, $|\det(\mathbf{A})| = \prod_{i=1}^n \sigma_i$.
- 6) Si \mathbf{A} es una matriz cuadrada y todos los valores singulares σ_i son diferentes, entonces los vectores singulares por la derecha y por la izquierda \mathbf{u}_i y \mathbf{v}_i son únicos, salvo signo.

3.4 Aplicación de la descomposición en valores singulares: compresión de imágenes

La compresión de imágenes que permite hacer la descomposición en valores singulares se basa en el resultado siguiente:

Toda matriz \mathbf{A} de dimensión $m \times n$ y rango r ($r < \min\{m, n\}$) se puede descomponer como suma de matrices r de rango 1:

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T,$$

donde σ_j son los valores singulares, \mathbf{u}_j son los vectores singulares por la izquierda y \mathbf{v}_j son los vectores singulares por la derecha, $j = 1, \dots, r$, en la descomposición en valores singulares (SVD).

La matriz \mathbf{A} representará una imagen en escala de grises rectangular de $m \times n$ píxeles. En general, la imagen contendrá píxeles que estarán relacionados entre sí, es decir, los píxeles cercanos tienen alta correlación. En un caso extremo, por ejemplo, si tenemos la fotografía de un campo de fútbol, muchos píxeles corresponderán a la zona del césped con tonalidades de verde muy similares. Por lo tanto, en general, el rango de la matriz r puede ser significativamente inferior al mínimo del número de filas o columnas de la matriz.

Ahora bien, la compresión en la imagen proviene del resultado siguiente:

$\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T$ es una matriz de dimensión $m \times n$ y rango r ($r < \min\{m, n\}$). Consideramos la ν -ésima suma parcial:

$$\mathbf{A}_\nu = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_\nu \mathbf{u}_\nu \mathbf{v}_\nu^T = \sum_{j=1}^{\nu} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

Así pues, la diferencia en norma 2 entre estas dos matrices es igual a:

$$\|\mathbf{A} - \mathbf{A}_\nu\|_2 = \sigma_{\nu+1}.$$

Norma 2 de una matriz

La norma 2 de una matriz corresponde a $\|\mathbf{A}\|_2 = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$ y se define, en cierta forma, como se define la norma 2 de un vector.

Sabemos que en la descomposición en valores singulares, los valores singulares:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$$

están ordenados de manera decreciente. Por lo tanto, si observamos que $\sigma_5 = 10^{-2}$, habrá que considerar la cuarta suma parcial:

$$\mathbf{A}_4 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \sigma_3 \mathbf{u}_3 \mathbf{v}_3^T + \sigma_4 \mathbf{u}_4 \mathbf{v}_4^T$$

para garantizar que la matriz original y la cuarta suma parcial tienen una diferencia, en norma 2, igual a $\sigma_5 = 10^{-2}$, que es un error pequeño.

Por lo tanto, la compresión de imágenes se basará en esta estrategia:

Consideramos que \mathbf{A} es la matriz que representa una imagen rectangular, en escala de grises, de $m \times n$ píxeles y que $\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T$ es la descomposición de la imagen almacenada en la matriz \mathbf{A} como suma de matrices r de rango 1. Entonces, la ν -ésima suma parcial

$$\mathbf{A}_\nu = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_\nu \mathbf{u}_\nu \mathbf{v}_\nu^T = \sum_{j=1}^{\nu} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

almacena la imagen comprimida con un error en norma 2 igual a $\sigma_{\nu+1}$.

Ejemplo

En este ejemplo tendremos en cuenta una fotografía del Parque Nacional de Yosemite (California, Estados Unidos de América), tal como se puede ver en la figura 7 en escala RGB y de medida 750×1.000 píxeles. La imagen se puede representar como una matriz de 750 filas y 1.000 columnas para cada color: rojo (R, *red*), verde (G, *green*) y azul (B, *blue*). Como la fotografía es en color, aplicaremos la descomposición en valores singulares a cada uno de los colores para volver a formar la imagen.

Figura 7. Imagen del Parque Nacional de Yosemite (en escala RGB y 750×1.000 píxeles)



Fuente: elaboración propia

Usaremos R para calcular la descomposición en valores singulares y las imágenes comprimidas. Primero cargaremos la librería `jpeg` y, después, leeremos la imagen `YOSEMITE.jpg`. La variable `yosemite` es una matriz tridimensional, donde para cada color tenemos 750 filas y 1.000 columnas. Podemos comprobar la dimensión de las matrices con las instrucciones `nrow(yosemite)` y `ncol(yosemite)`.

```
library(jpeg)
yosemite<-readJPEG('YOSEMITE.jpg')
nrow(yosemite)
ncol(yosemite)
```

A continuación generamos las tres matrices a las que aplicaremos la descomposición en valores singulares por separado. La matriz `r` contendrá el color rojo; la matriz `g`, el verde, y la matriz `b`, el azul. Fijaos en que hemos accedido a la dimensión del color mediante la instrucción `yosemite[,1]` (para el color rojo), `yosemite[,2]` (para el color verde) o `yosemite[,3]` para el color azul. Como `yosemite` era una matriz de dimensión $750 \times 1.000 \times 3$, hemos accedido a la tercera componente de la matriz mencionada.

```
r <- yosemite[,1]
g <- yosemite[,2]
b <- yosemite[,3]
```

Después calculamos la descomposición en valores singulares mediante la instrucción `svd`. Almacenamos los resultados en estas variables: `yosemite.r.svd`, `yosemite.g.svd` y `yosemite.b.svd`. Juntamos las tres descomposiciones en una lista mediante la instrucción `list`.

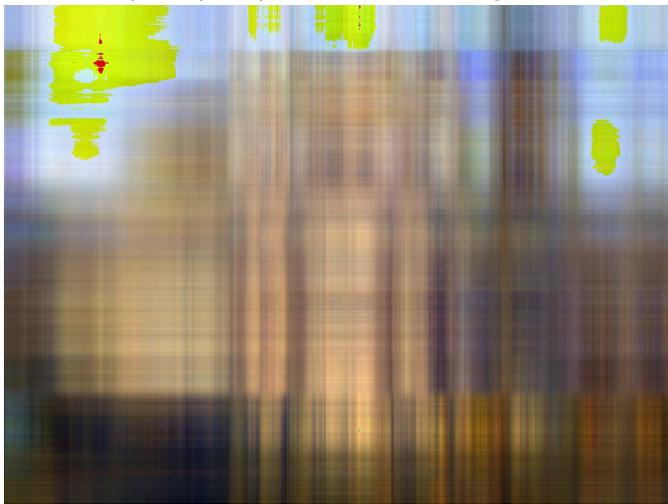
```
yosemite.r.svd <- svd(r)
yosemite.g.svd <- svd(g)
yosemite.b.svd <- svd(b)
rgb.svds <- list(yosemite.r.svd, yosemite.g.svd, yosemite.b.svd)
```

Cada variable (`yosemite.r.svd`, `yosemite.g.svd` y `yosemite.b.svd`) es una estructura de datos que contiene tres campos: `u` (matriz que contiene los vectores singulares por la izquierda), `d` (vector con los valores singulares) y `v` (matriz que contiene los vectores singulares por la derecha). Accedemos a cada uno de estos campos escribiendo, por ejemplo, `yosemite.r.svd$u` o `yosemite.g.svd$d`.

Para calcular de una manera rápida y compacta la ν -ésima suma parcial usamos la instrucción `sapply`, que permite aplicar una función a una lista. En un primer ejemplo, si fijamos $\nu = 3$, la tercera suma parcial se puede almacenar en la variable `a`. Finalmente, se puede guardar la imagen comprimida resultante usando la instrucción `writeJPEG`. La imagen que resulta se puede ver en la figura 8.

```
nu <- 3
a <- sapply(rgb.svds, function(i) {
  yosemite.compress <- i$u[,1:nu]
  %*% diag(i$d[1:nu]) %*% t(i$v[,1:nu])},
  simplify = 'array')
writeJPEG(a, 'yose003.jpg')
```

Figura 8. Imagen del Parque Nacional de Yosemite (en escala RGB, 750×1.000 píxeles y comprimida a tres valores singulares)



Fuente: elaboración propia

Del mismo modo podemos calcular la imagen comprimida que resulta de la ν -ésima suma parcial si $\nu = 20$, que hemos representado en la figura 9.

```
nu <- 20
a <- sapply(rgb.svds, function(i) {
  yosemite.compress <- i$u[,1:nu]
  %*% diag(i$d[1:nu]) %*% t(i$v[,1:nu])},
  simplify = 'array')
writeJPEG(a, 'yose020.jpg')
```

Finalmente, la imagen comprimida que resulta de la ν -ésima suma parcial si $\nu = 200$ se puede ver en la figura 10.

```
nu <- 200
a <- sapply(rgb.svds, function(i) {
  yosemite.compress <- i$u[,1:nu]
  %*% diag(i$d[1:nu]) %*% t(i$v[,1:nu])},
  simplify = 'array')
writeJPEG(a, 'yose200.jpg')
```

A modo de ejemplo y por lo que respecta a la imagen de la figura 10 comprimida a doscientos valores singulares, los valores singulares $\sigma_{201}^r, \sigma_{201}^g$ y σ_{201}^b , correspondientes a las matrices de colores rojo, verde y azul, respectivamente, son:

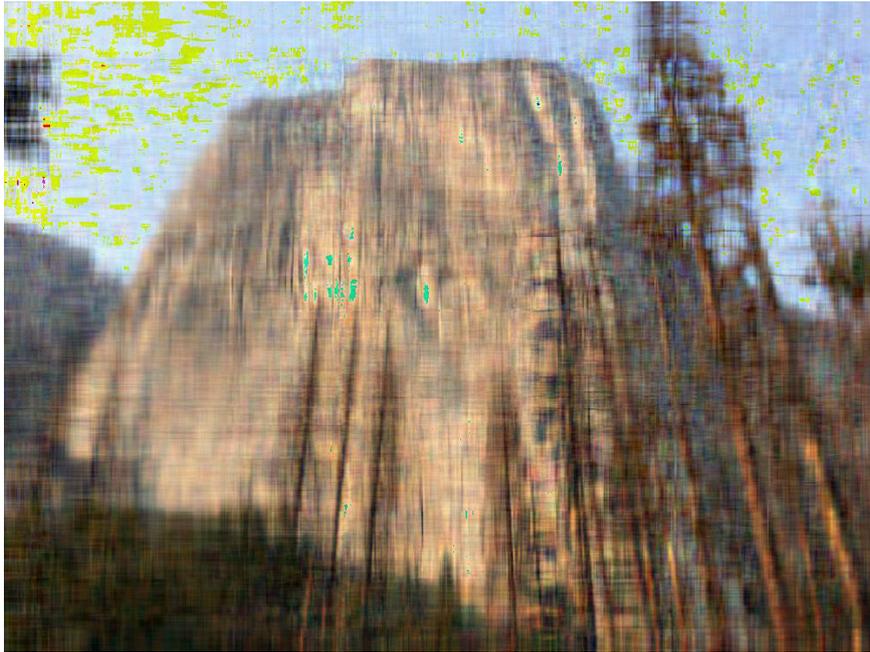
$$\sigma_{201}^r = 4.195227,$$

$$\sigma_{201}^g = 4.19045,$$

$$\sigma_{201}^b = 4.033507.$$

Hemos obtenido estos valores con `yosemite.r.svd$d[201]`, `yosemite.g.svd$d[201]` y `yosemite.b.svd$d[201]`, respectivamente.

Figura 9. Imagen del Parque Nacional de Yosemite (en escala RGB, 750×1.000 píxeles y comprimida a veinte valores singulares)



Fuente: elaboración propia

¿Cómo debemos interpretar estas cantidades? Recordad que nuestra imagen inicial está formada por tres matrices: una para el color rojo, una para el color verde y otra para el color azul. Es decir, tenemos tres matrices, \mathbf{A}^r , \mathbf{A}^g y \mathbf{A}^b , que hemos aproximado (en la figura 10, por una imagen comprimida formada por sumas parciales) para cada uno de los colores: \mathbf{A}_{200}^r , \mathbf{A}_{200}^g y \mathbf{A}_{200}^b . Entonces, el error para cada uno de los colores, en norma 2, es:

$$\|\mathbf{A}^r - \mathbf{A}_{200}^r\| = \sigma_{201}^r = 4.195227$$

$$\|\mathbf{A}^g - \mathbf{A}_{200}^g\| = \sigma_{201}^g = 4.19045$$

$$\|\mathbf{A}^b - \mathbf{A}_{200}^b\| = \sigma_{201}^b = 4.033507$$

El error en la compresión de la imagen es pequeño, pese a que aún es observable. La razón es que la imagen presenta un número importante de contrastes y los valores singulares decrecen poco a poco. En términos de espacio necesario para almacenar la imagen original hacen falta

$$750 \times 1.000 = 750.000$$

elementos. Si usamos la suma parcial con $\nu = 200$, hay que almacenar 200 matrices de rango 1 definidas por $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$, donde \mathbf{u}_i tiene dimensión 750 y \mathbf{v}_i tiene dimensión 1.000. Esto corresponde a:

$$\nu \cdot (m + n) + \nu = 200 \cdot (750 + 1.000) + 200 = 350.200$$

elementos. En este caso, esto representa un:

$$\frac{350.200}{750.000} \times 100 \% = 46.70 \%$$

del tamaño original.

Figura 10. Imagen del Parque Nacional de Yosemite (en escala RGB, 750×1.000 píxeles y comprimida a 200 valores singulares)



Fuente: elaboración propia

Ejemplo

Este ejemplo es igual que el anterior, en el sentido de que consideramos una imagen en escala de colores RGB y de dimensión 750×1.000 píxeles. Sin embargo, la imagen es un poco más sencilla, como se puede ver en la figura 11.

Figura 11. Imagen de una flor de hibisco sobre una tela (en escala RGB y 750×1.000 píxeles)



Fuente: elaboración propia

En la figura 12 hemos representado la imagen comprimida con los doscientos primeros valores singulares. Los valores singulares $\sigma_{201}^r, \sigma_{201}^g$ y σ_{201}^b , correspondientes a las matrices de colores rojo, verde y azul, respectivamente, son, en este caso:

$$\begin{aligned}\sigma_{201}^r &= 3.1221, \\ \sigma_{201}^g &= 3.483641, \\ \sigma_{201}^b &= 3.200724\end{aligned}$$

Figura 12. Imagen de una flor de hibisco sobre una tela (en escala RGB, 750×1.000 píxeles y comprimida a 200 valores singulares)



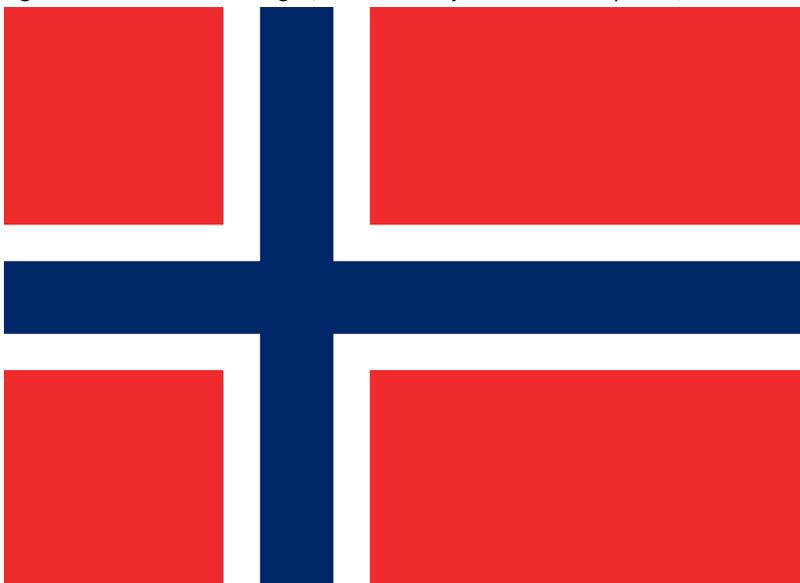
Fuente: elaboración propia

Estos valores son inferiores a los valores singulares del ejemplo anterior. Esto significa que con el mismo número de valores singulares (200) obtenemos una compresión más buena. De hecho, esta mejora es perfectamente observable.

Ejemplo

Finalmente, veremos un ejemplo de compresión extrema a partir de la imagen de la bandera de Noruega. Se trata de una imagen en escala de colores RGB, de dimensión 1.090×1.500 píxeles, tal como muestra la figura 13.

Figura 13. Bandera de Noruega (en escala RGB y 1.090×1.500 píxeles)

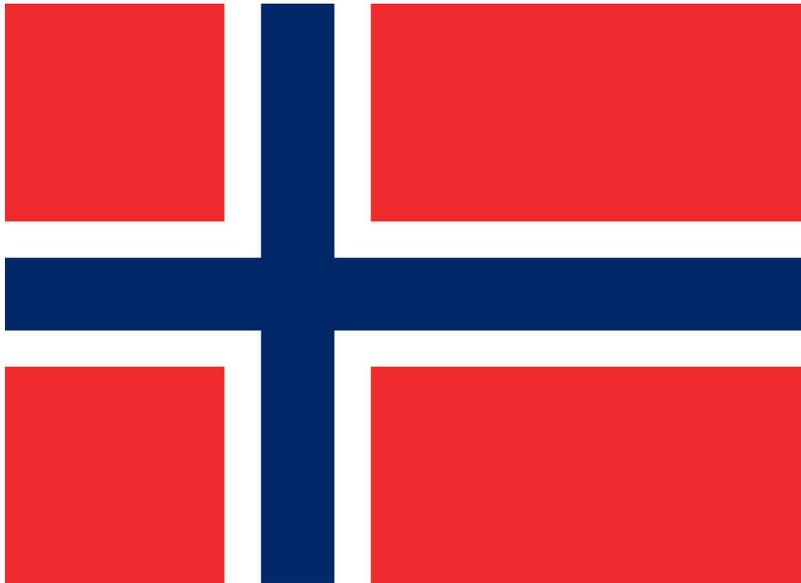


Fuente: <https://www.countryflags.com/en/norway-flag-image.html>

En la figura 14 hemos representado la imagen comprimida con los tres primeros valores singulares. Los valores singulares σ_4^r, σ_4^g y σ_4^b , correspondientes a las matrices de colores rojo, verde y azul, respectivamente, son, en este caso:

$$\begin{aligned}\sigma_4^r &= 5.199605 \cdot 10^{-2}, \\ \sigma_4^g &= 3.652678 \cdot 10^{-2}, \\ \sigma_4^b &= 5.095715 \cdot 10^{-2}.\end{aligned}$$

Figura 14. Bandera de Noruega (en escala RGB, 1.090×1.500 píxeles y comprimida a tres valores singulares)



Fuente: elaboración propia

El error en la compresión de la imagen es inapreciable. En términos de espacio necesario para almacenar la imagen original hacen falta

$$1.090 \times 1.500 = 1.635.000$$

elementos. Si usamos la suma parcial con $\nu = 3$, hay que almacenar tres matrices de rango 1 definidas por $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$, donde \mathbf{u}_i tiene dimensión 1.090 y \mathbf{v}_i tiene dimensión 1.500. Esto corresponde a

$$\nu \cdot (m + n) + \nu = 3 \cdot (1.090 + 1.500) + 3 = 6.423$$

elementos. En este caso, esto representa un:

$$\frac{6.423}{1.635.000} \times 100 \% = 0.39 \%$$

del tamaño original.

Resumen

En este módulo hemos visto dos técnicas, el análisis de componentes principales (PCA) y la descomposición en valores singulares (SVD), donde la base matemática en ambos casos es el cálculo de valores y vectores propios.

En lo que respecta al análisis de componentes principales, hay que tener en cuenta algunos puntos importantes en esta estrategia:

- 1) Los datos que medimos se pueden organizar, habitualmente, en forma de matriz, donde el número de columnas (m) representa las diferentes variables que medimos y el número de filas (n) representa los elementos de la muestra.
- 2) Hay que normalizar los datos para evitar los efectos que las magnitudes de cada variable tendrían sobre el análisis. Así pues, debemos restar por la media de cada variable y dividir por la desviación típica de cada variable.
- 3) Se calcula la matriz de covarianzas por medio del producto de matrices $\frac{1}{n-1} \mathbf{X}^T \mathbf{X}$. El resultado es una matriz simétrica y definida positiva, donde los elementos de la diagonal son todos unos.
- 4) Se calculan los valores y vectores propios de la matriz de covarianzas y se ordenan de grande a pequeño. El vector propio asociado al valor propio más grande se denomina *primera componente principal*. El vector propio asociado al segundo valor propio más grande se denomina *segunda componente principal*. Y así sucesivamente. En general, los vectores propios de la matriz de covarianzas son las componentes principales.
- 5) La proyección de los datos originales sobre el espacio vectorial generado por las componentes principales permite descubrir características que pueden pasar desapercibidas. Esta proyección también permite reducir la dimensionalidad de los datos originales.

En relación con la descomposición en valores singulares, los elementos clave en esta estrategia son:

- 1) La descomposición en valores singulares descompone una matriz \mathbf{A} de dimensión $n \times m$ como producto de tres matrices, $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, donde \mathbf{U}, \mathbf{V} son matrices unitarias y $\mathbf{\Sigma}$ es una matriz diagonal (en la versión reducida). Los elementos de la diagonal se denominan *valores singulares*, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, donde r es el rango de la matriz \mathbf{A} .
- 2) El cuadrado de los valores singulares $\sigma_1^2, \dots, \sigma_r^2$ son los valores propios no nulos tanto de la matriz $\mathbf{A}\mathbf{A}^T$ como de la matriz $\mathbf{A}^T\mathbf{A}$.

- 3) Las columnas ortonormales de la matriz \mathbf{U} son los vectores propios de la matriz $\mathbf{A}\mathbf{A}^T$ y reciben el nombre de *vectores singulares por la izquierda*.
- 4) Las columnas ortonormales de la matriz \mathbf{V} son los vectores propios de la matriz $\mathbf{A}^T\mathbf{A}$ y reciben el nombre de *vectores singulares por la derecha*.
- 5) Los vectores singulares por la izquierda y los vectores singulares por la derecha son dos bases que diagonalizan la matriz \mathbf{A} . Es decir, $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$, $i \leq r$. De forma equivalente, $\mathbf{A}\mathbf{V} = \Sigma\mathbf{U}$.
- 6) Una de las aplicaciones más interesantes de la descomposición en valores singulares es la compresión de imágenes.

Ejercicios de autoevaluación

1. Considerad esta matriz, que contiene el valor de dos variables para una muestra de cinco elementos:

$$\mathbf{X}_0 = \begin{bmatrix} 5 & -1 \\ 4 & 1 \\ 3 & 0 \\ 2 & 1 \\ 0 & -1 \end{bmatrix}$$

Calculad la media aritmética de cada columna para obtener la matriz centrada \mathbf{X} . Calculad la matriz de covarianzas $\mathbf{C}_\mathbf{X}$ y los valores propios λ_1 y λ_2 . ¿Cuál es la dirección de la primera componente principal?

2. Considerad esta matriz, que contiene el valor de dos variables para una muestra de seis elementos:

$$\mathbf{X}_0 = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \\ 3 & 0 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}$$

Calculad la media aritmética de cada columna para obtener la matriz centrada \mathbf{X} . Calculad la matriz de covarianzas $\mathbf{C}_\mathbf{X}$ y los valores propios λ_1 y λ_2 . ¿Cuál es la dirección de la primera componente principal?

3. Considerad esta matriz, que contiene el valor de tres variables para una muestra de cuatro elementos:

$$\mathbf{X}_0 = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \\ 0 & 2 & -1 \\ 0 & -2 & -1 \end{bmatrix}$$

Calculad la media aritmética y la desviación típica de cada columna para obtener la matriz normalizada \mathbf{X} . Calculad la matriz de covarianzas $\mathbf{C}_\mathbf{X}$ y los valores propios λ_1, λ_2 y λ_3 . ¿Cuál es la dirección de la primera componente principal?

4. Considerad esta matriz:

$$\mathbf{A} = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix}$$

Calculad los valores propios de la matriz \mathbf{A} . Encontrad los valores singulares de la matriz $\mathbf{A}^T \mathbf{A}$. Calculad \mathbf{V} mediante los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$ y \mathbf{U} mediante los vectores propios de la matriz $\mathbf{A} \mathbf{A}^T$. Comprobad que $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$.

5. Considerad esta matriz:

$$\mathbf{A} = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$$

Calculad los valores propios de la matriz \mathbf{A} . Encontrad los valores singulares de la matriz $\mathbf{A}^T \mathbf{A}$. Calculad \mathbf{V} mediante los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$ y \mathbf{U} mediante los vectores propios de la matriz $\mathbf{A} \mathbf{A}^T$. Comprobad que $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$.

6. Considerad esta matriz:

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

Calculad $\mathbf{A}^T \mathbf{A}$, \mathbf{V} , $\mathbf{\Sigma}$ y $\mathbf{u}_i = \mathbf{A} \mathbf{v}_i / \sigma_i$ y la descomposición en valores singulares (SVD) completa. Comprobad que $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$.

7. Considerad una matriz \mathbf{A} cualquiera. Demostrad que los valores singulares de esta matriz σ_i son únicos.

8. Considerad una matriz \mathbf{A} cualquiera. Demostrad que el rango de la matriz coincide con el número de valores singulares no nulos.

9. Considerad una matriz cuadrada \mathbf{A} cualquiera. Demostrad que el valor absoluto del determinante de esta matriz es igual que el producto de los valores singulares, es decir:

$$|\det(\mathbf{A})| = \prod_{i=1}^n \sigma_i$$

Solucionario

1. La media aritmética de la primera columna es 2.8 y la de la segunda, 0. Por lo tanto, la matriz centrada es:

$$\mathbf{X} = \begin{bmatrix} 2.2 & -1 \\ 1.2 & 1 \\ 0.2 & 0 \\ -0.8 & 1 \\ -2.8 & -1 \end{bmatrix}$$

La matriz de covarianzas es:

$$\mathbf{C}_X = \frac{1}{4} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 3.70 & 0.25 \\ 0.25 & 1.00 \end{bmatrix}$$

Observad que la matriz de covarianzas no tiene unos en la diagonal porque no hemos dividido las columnas por las desviaciones típicas.

Los valores propios de la matriz de covarianzas son:

$$\lambda_1 = 3.722953$$

$$\lambda_2 = 0.977047$$

Fijaos en que

$$\lambda_1 + \lambda_2 = 4.7 = \text{tr}(\mathbf{C}_X).$$

Finalmente, la dirección de la primera componente principal es

$$p_1^T = [-0.99581173, -0.09142755].$$

La resolución en \mathbb{R} de este problema se puede ver en la figura 15.

Figura 15

```
> a<-c(5,-1,4,1,3,0,2,1,0,-1)
> X0<-matrix(a,ncol=2,nrow=5,byrow=TRUE)
> X0[,1]<-X0[,1]-mean(X0[,1])
> X0[,2]<-X0[,2]-mean(X0[,2])
> X<-X0
> X
      [,1] [,2]
[1,]  2.2  -1
[2,]  1.2   1
[3,]  0.2   0
[4,] -0.8   1
[5,] -2.8  -1
> CX<-t(X)%*X/4
> CX
      [,1] [,2]
[1,]  3.70  0.25
[2,]  0.25  1.00
> eigCX<-eigen(CX)
> eigCX$values
[1]  3.722953  0.977047
> eigCX$vectors
      [,1] [,2]
[1,] -0.99581173  0.09142755
[2,] -0.09142755 -0.99581173
```

Fuente: elaboración propia

2. La media aritmética de la primera columna es 2 y la de la segunda, 0.5. Por lo tanto, la matriz centrada es:

$$\mathbf{X} = \begin{bmatrix} -1 & 0.5 \\ 0 & -0.5 \\ 1 & 0.5 \\ 1 & -0.5 \\ 0 & 0.5 \\ -1 & -0.5 \end{bmatrix}$$

La matriz de covarianzas es:

$$\mathbf{C}_X = \frac{1}{5} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.3 \end{bmatrix}$$

Observad que la matriz de covarianzas no tiene unos en la diagonal porque no hemos dividido las columnas por las desviaciones típicas. Fijaos en que la matriz de covarianzas es, en este caso, una matriz diagonal.

Los valores propios de la matriz de covarianzas son:

$$\lambda_1 = 0.8$$

$$\lambda_2 = 0.3$$

Fijaos en que

$$\lambda_1 + \lambda_2 = 1.1 = \text{tr}(\mathbf{C}_X).$$

Finalmente, la dirección de la primera componente principal es

$$p_1^T = [-1, 0].$$

Observad que, como la matriz de covarianzas es diagonal, la primera componente principal es igual (en lo que respecta a la dirección) que la variable que hemos considerado en la primera columna de la matriz \mathbf{X}_0 .

La resolución en \mathbb{R} de este problema se puede ver en la figura 16.

Figura 16

```
> a<-c(1,2,3,3,2,1,1,0,1,0,1,0)
> X0<-matrix(a,ncol=2,nrow=6,byrow=FALSE)
> X0[,1]<-X0[,1]-mean(X0[,1])
> X0[,2]<-X0[,2]-mean(X0[,2])
> X<-X0
> X
      [,1] [,2]
[1,]  -1  0.5
[2,]   0 -0.5
[3,]   1  0.5
[4,]   1 -0.5
[5,]   0  0.5
[6,]  -1 -0.5
> CX<-t(X)%*%X/5
> CX
      [,1] [,2]
[1,]  0.8  0.0
[2,]  0.0  0.3
> eigCX<-eigen(CX)
> eigCX$values
[1] 0.8 0.3
> eigCX$vectors
      [,1] [,2]
[1,]  -1  0
[2,]   0 -1
```

Fuente: elaboración propia

3. La media aritmética de la primera columna es 0; la de la segunda, 0, y la de la tercera, 0. Las desviaciones típicas de las columnas son 0.8164966, 1.632993 y 1.154701, respectivamente. Por lo tanto, la matriz escalonada es:

$$\mathbf{X} = \begin{bmatrix} 1.224745 & 0.000000 & 0.8660254 \\ -1.224745 & 0.000000 & 0.8660254 \\ 0.000000 & 1.224745 & -0.8660254 \\ 0.000000 & -1.224745 & -0.8660254 \end{bmatrix}$$

La matriz de covarianzas es:

$$\mathbf{C}_X = \frac{1}{6} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Observad que ahora la matriz de covarianzas sí que tiene unos en la diagonal, ya que hemos dividido las columnas por las desviaciones típicas. Fijaos en que la matriz de covarianzas es, en este caso, una matriz diagonal.

Los valores propios de la matriz de covarianzas son:

$$\lambda_1 = 1$$

$$\lambda_2 = 1$$

$$\lambda_3 = 1$$

Fijaos en que

$$\lambda_1 + \lambda_2 + \lambda_3 = 3 = \text{tr}(\mathbf{C}_X).$$

Finalmente, la dirección de la primera componente principal es

$$p_1^T = [0,0,1].$$

En este caso, después de haber obtenido los tres valores propios iguales a 1, no hay ningún valor más grande que el resto. La segunda y tercera componentes principales son, según los resultados que obtenemos en \mathbb{R} :

$$p_2^T = [0,1,0]$$

$$p_3^T = [1,0,0]$$

Por tanto, teniendo en cuenta que la matriz de covarianza es la matriz identidad, habría sido más normal definir las componentes principales así:

$$p_1^T = [1,0,0]$$

$$p_2^T = [0,1,0]$$

$$p_3^T = [0,0,1]$$

La resolución en \mathbb{R} de este problema se puede ver en la figura 17.

Figura 17

```
> a<-c(1,-1,0,0,0,0,2,-2,1,1,-1,-1)
> X0<-matrix(a,ncol=3,nrow=4,byrow=FALSE)
> X0[,1]<-(X0[,1]-mean(X0[,1]))/sd(X0[,1])
> X0[,2]<-(X0[,2]-mean(X0[,2]))/sd(X0[,2])
> X0[,3]<-(X0[,3]-mean(X0[,3]))/sd(X0[,3])
> X<-X0
> X
      [,1] [,2] [,3]
[1,] 1.224745 0.000000 0.8660254
[2,] -1.224745 0.000000 0.8660254
[3,] 0.000000 1.224745 -0.8660254
[4,] 0.000000 -1.224745 -0.8660254
> CX<-t(X)%*%X/3
> CX
      [,1] [,2] [,3]
[1,] 1 0 0
[2,] 0 1 0
[3,] 0 0 1
> eigCX<-eigen(CX)
> eigCX$values
[1] 1 1 1
> eigCX$vectors
      [,1] [,2] [,3]
[1,] 0 0 1
[2,] 0 1 0
[3,] 1 0 0
```

Fuente: elaboración propia

4. El determinante de la matriz \mathbf{A} es 0 y la traza, también. Por lo tanto, sabemos que los valores propios de la matriz \mathbf{A} satisfacen:

$$\lambda_1 + \lambda_2 = \text{tr}(\mathbf{A}) = 0$$

$$\lambda_1 \lambda_2 = \det(\mathbf{A}) = 0$$

Esto implica que $\lambda_1 = \lambda_2 = 0$. Fijaos en que el rango de la matriz \mathbf{A} es $r = 1$. Por otro lado, la matriz $\mathbf{A}^T \mathbf{A}$ es igual a:

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 16 \end{bmatrix},$$

que tiene como valores propios $\lambda_1 = 16$ y $\lambda_2 = 0$. Por tanto, el único valor singular es $\sigma_1 = \sqrt{\lambda_1} = 4$. Los vectores propios de la matriz $\mathbf{A}^T \mathbf{A}$ son:

$$\mathbf{v}_1^T = [0, 1]$$

$$\mathbf{v}_2^T = [-1, 0],$$

de forma que la matriz \mathbf{V} es igual a:

$$\mathbf{V} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Por otro lado, la matriz $\mathbf{A}\mathbf{A}^T$ es igual a:

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix},$$

que tiene como valores propios $\lambda_1 = 16$ y $\lambda_2 = 0$. Por lo tanto, el único valor singular es $\sigma_1 = \sqrt{\lambda_1} = 4$. Los vectores propios de la matriz $\mathbf{A}\mathbf{A}^T$ son:

$$\mathbf{u}_1^T = [1, 0]$$

$$\mathbf{u}_2^T = [0, 1],$$

de forma que la matriz \mathbf{V} es igual a:

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Se puede comprobar, pues, que $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$:

$$\begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^T$$

5. El determinante de la matriz \mathbf{A} es -4 y la traza es 0 . Por tanto, sabemos que los valores propios de la matriz \mathbf{A} satisfacen:

$$\lambda_1 + \lambda_2 = \text{tr}(\mathbf{A}) = 0$$

$$\lambda_1 \lambda_2 = \det(\mathbf{A}) = -4$$

Esto implica que $\lambda_1 = 2$ y $\lambda_2 = -\lambda_1 = -2$. Fijaos en que el rango de la matriz \mathbf{A} es $r = 2$. Por otro lado, la matriz $\mathbf{A}^T\mathbf{A}$ es igual a:

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix},$$

que tiene como valores propios $\lambda_1 = 16$ y $\lambda_2 = 1$. Por tanto, los valores singulares son $\sigma_1 = \sqrt{\lambda_1} = 4$ y $\sigma_2 = \sqrt{\lambda_2} = 1$. Los vectores propios de la matriz $\mathbf{A}^T\mathbf{A}$ son:

$$\mathbf{v}_1^T = [0, -1]$$

$$\mathbf{v}_2^T = [-1, 0],$$

de forma que la matriz \mathbf{V} es igual a:

$$\mathbf{V} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Por lo tanto, la matriz $\mathbf{A}\mathbf{A}^T$ es igual a:

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix}$$

que tiene como valores propios $\lambda_1 = 16$ y $\lambda_2 = 1$. Por tanto, los valores singulares son $\sigma_1 = \sqrt{\lambda_1} = 4$ y $\sigma_2 = \sqrt{\lambda_2} = 1$. Los vectores propios de la matriz $\mathbf{A}\mathbf{A}^T$ son:

$$\mathbf{u}_1^T = [-1, 0]$$

$$\mathbf{u}_2^T = [0, -1],$$

de forma que la matriz \mathbf{V} es igual a:

$$\mathbf{V} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

Se puede comprobar, pues, que $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$:

$$\begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}^T$$

6. La matriz $\mathbf{A}^T\mathbf{A}$ es igual a:

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix},$$

que tiene como valores propios $\lambda_1 = 8$ y $\lambda_2 = 2$. Por tanto, los valores singulares son $\sigma_1 = \sqrt{\lambda_1} = 2\sqrt{2}$ y $\sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$. Los vectores propios de la matriz $\mathbf{A}^T\mathbf{A}$ son:

$$\mathbf{v}_1^T = [-\sqrt{2}/2, -\sqrt{2}/2]$$

$$\mathbf{v}_2^T = [-\sqrt{2}/2, \sqrt{2}/2],$$

de forma que la matriz \mathbf{V} es igual a:

$$\mathbf{V} = \begin{bmatrix} -\sqrt{2}/2 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

La matriz Σ es igual a:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

Calcularemos los vectores \mathbf{u}_i siguiendo la indicación $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i/\sigma_i$. En efecto:

$$\mathbf{u}_1 = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Se puede comprobar, pues, que $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$:

$$\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} -\sqrt{2}/2 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}^T$$

7. Los valores singulares de la matriz \mathbf{A} son las raíces cuadradas de los valores propios de la matriz $\mathbf{A}^T\mathbf{A}$. Sabemos que los valores propios de cualquier matriz son únicos, lo que implica que los valores singulares de la matriz \mathbf{A} son también únicos.

8. Consideramos la descomposición en valores singulares completa de la matriz \mathbf{A} . Esto implica la existencia de dos matrices, \mathbf{U} y \mathbf{V} , de rango máximo, ya que están formadas por una base de vectores ortonormales de los correspondientes espacios vectoriales. Entonces:

$$\text{rango}(\mathbf{A}) = \text{rango}(\mathbf{U}\Sigma\mathbf{V}^T) = \text{rango}(\Sigma)$$

El rango de la matriz diagonal Σ es igual que el número de valores no nulos de su diagonal, es decir, es el número de valores singulares no nulos de la matriz \mathbf{A} .

9. Consideramos la descomposición en valores singulares completa de la matriz \mathbf{A} . Esto implica la existencia de dos matrices, \mathbf{U} y \mathbf{V} , de rango máximo, ya que están formadas por una base de vectores ortonormales de los correspondientes espacios vectoriales. Teniendo en cuenta que el determinante de un producto de matrices es igual que el producto de los determinantes, y teniendo en cuenta también que el determinante de una matriz ortogonal es 1, obtenemos:

$$|\det(\mathbf{A})| = |\det(\mathbf{U}\Sigma\mathbf{V}^T)| = |\det(\mathbf{U}) \det(\Sigma) \det(\mathbf{V}^T)| = |\det(\Sigma)| = \prod_{i=1}^n \sigma_i$$

Glosario

componentes principales de una matriz f Vectores propios de la matriz de covarianzas correspondiente.

matriz de covarianzas f Matriz cuadrada que en la posición (i,j) contiene la covarianza de las variables i y j .

matriz unitaria f Matriz en la que el producto escalar de dos columnas diferentes es igual a cero y en la que el producto escalar de dos columnas iguales equivale a 1. La inversa de una matriz unitaria es igual que su transpuesta.

primera componente principal f Componente principal asociada al valor propio máximo de la matriz de covarianzas.

valores singulares m Raíces cuadradas de los valores propios de las matrices $\mathbf{A}\mathbf{A}^T$ y $\mathbf{A}^T\mathbf{A}$.

vectores singulares por la izquierda m Vectores propios de la matriz $\mathbf{A}\mathbf{A}^T$.

vectores singulares por la derecha m Vectores propios de la matriz $\mathbf{A}^T\mathbf{A}$.

Bibliografía

Castellet, Manuel; Llerena, Irene (2010). *Álgebra lineal y geometría*. Barcelona: Editorial Reverté.

Jolliffe, Ian T. (2002). *Principal Component Analysis*. Berlín: Springer.

Liesen, Jörg; Mehrmann, Volker (2015). *Linear Algebra*. Berlín: Springer.

Peña, Daniel (2013). *Análisis de datos multivariantes*. Madrid: McGraw-Hill España.

Strang, Gilbert (2016). *Introduction to Linear Algebra*. Wellesley (Massachusetts, EE. UU.): Wellesley-Cambridge Press.