Unidad 1: Reducción de dimensionalidad

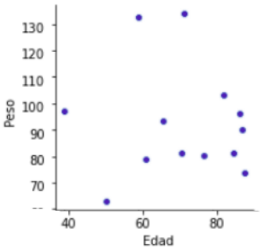
Vectores y valores propios de matrices simétricas

Blanca Vázquez

11 de agosto de 2022

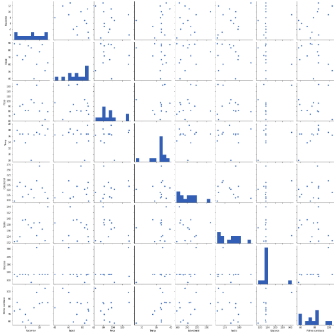
Introducción

Paciente	Edad	Peso	Temp	Colesterol	Sodio	Glucosa	Ritmo cardiaco	…
1	88.78	74.00	34.40	96.00	132.33	137.35	73.77	93.19
2	72.33	134.00	37.90	168.35	132.50	124.00	85.25	81.18
3	88.27	90.00	36.71	188.00	134.00	137.35	75.85	81.18
4	66.68	93.19	36.71	116.00	139.50	137.35	58.82	103.00
5	71.70	81.18	33.77	168.35	139.67	137.35	69.30	97.00
6	85.92	81.18	36.71	128.90	138.00	134.00	61.48	132.50
7	83.09	103.00	28.03	153.25	137.00	85.00	57.28	80.20
8	39.91	97.00	36.40	112.00	138.60	137.35	78.08	93.19
9	60.15	132.50	37.20	196.00	134.33	309.00	98.81	81.18
10	77.69	80.20	36.71	269.00	138.71	137.35	83.79	81.18
11	87.45	96.20	39.60	160.00	136.50	137.35	84.91	103.00
12	51.26	63.10	36.30	138.60	132.67	137.35	104.43	97.00
13	62.16	78.90	38.38	109.00	144.00	137.35	85.08	132.50



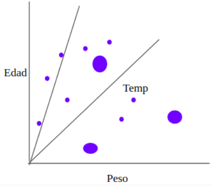
Paciente	Edad	Peso
1	88.78	74.00
2	72.33	134.00
3	88.27	90.00
4	66.68	93.19
5	71.70	81.18
6	85.92	81.18
7	83.09	103.00
8	39.91	97.00
9	60.15	132.50
10	77.69	80.20
11	87.45	96.20
12	51.26	63.10
13	62.16	78.90

Una base de datos, con k número de atributos, es posible visualizar los datos como una nube de puntos en un espacio de k-dimensiones.



¿Qué pasa con 4, 5 o más variables?

Paciente	Edad	Peso	Temp	Colesterol	Sodio	Glucosa	Ritmo cardiaco	...
1	88.78	74.00	34.40	96.00	132.33	137.35	73.77	93.19
2	72.33	134.00	37.90	168.35	132.50	124.00	85.25	81.18
3	88.27	90.00	36.71	188.00	134.00	137.35	75.85	81.18
4	66.68	93.19	36.71	116.00	139.50	137.35	58.82	103.00
5	71.70	81.18	33.77	168.35	139.67	137.35	69.30	97.00
6	85.92	81.18	36.71	128.90	138.00	134.00	61.48	132.50
7	83.09	103.00	28.03	153.25	137.00	85.00	57.28	80.20
8	39.91	97.00	36.40	112.00	138.60	137.35	78.08	93.19
9	60.15	132.50	37.20	196.00	134.33	309.00	98.81	81.18
10	77.69	80.20	36.71	269.00	138.71	137.35	83.79	81.18
11	87.45	96.20	39.60	160.00	136.50	137.35	84.91	103.00
12	51.26	63.10	36.30	138.60	132.67	137.35	104.43	97.00
13	62.16	78.90	38.38	109.00	144.00	137.35	85.08	132.50



Tener un largo número de dimensiones puede ocasionar que tengamos datos no representativos.

Importancia de la reducción de la dimensionalidad

La reducción de la dimensionalidad consiste en **disminuir** el número de dimensiones en un espacio de características.

- Remover las características redundantes
- Menos poder de cómputo durante el entrenamiento
- Es posible reducir el sobreajuste
- Mejora el rendimiento
- Nos ayuda a reducir el espacio de almacenamiento

Las técnicas de reducción de dimensionalidad buscan **proyectar los datos a un sub-espacio de menor dimensión** que captura la esencia original de los datos.

- Métodos de proyección (álgebra lineal)
    - Análisis de componentes principales (PCA)
    - Análisis discriminante lineal (LDA)
    - Análisis discriminante generalizado (GDA)

Descomposición propia de una matriz

Una matriz de tamaño *nxn* puede descomponerse en un conjunto de vectores propios **(eigenvectores)** y valores propios **(eigenvalores)**.

Dada una matriz cuadrada M, λ una constante y v un vector columna distinto de cero con el mismo número de filas que M, λ es un valor propio de M y v su correspondiente vector propio de M, si $Mv = \lambda v$

$$Mv = \lambda v$$

Matriz cuadrada   Vector propio   Valor propio

- Dada la matriz cuadrada M, calcular los valores propios y los vectores propios

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

DEMOSTRAR QUE $Mv = \lambda v$

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

Uno de los vectores propios de M es:

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

y su correspondiente valor propio es 7.

Demostrar que $Mv = \lambda v$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Del lado izquierdo, multiplicamos M por v y del lado derecho se multiplica λ por v

$$\begin{bmatrix} 3/\sqrt{5} + 4/\sqrt{5} \\ 2/\sqrt{5} + 12/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

Finalmente, sumamos

$$\begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

¿Cómo se calculan los valores propios y vectores propios?

Recordemos la ecuación:

$$Mv = \lambda v$$

Empezamos con el lado derecho de la ecuación:

$$\lambda v$$

Buscamos tener una matriz que tenga en la diagonal λ y ceros en el resto de posiciones.

Añadir una matriz identidad

Una matriz identidad es una matriz de nxn que tiene 1s en la diagonal y 0s en el resto de posiciones

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Deseamos obtener esto:

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

Por lo tanto, obtenemos: $(\lambda I)v$

Añadir una matriz identidad

$$Mv = \lambda v$$
$$Mv = (\lambda I)v$$
$$Mv - (\lambda I)v = 0$$
$$M - \lambda I = 0$$

Importante, buscamos que el vector propio v sea distinto de 0

Sea v un vector propio que no es 0, entonces para resolver λ se debe calcular el determinante de:

$$|M - (\lambda I)| = 0$$

¿Determinante?

- El determinante de A $|A|$: es el número obtenido al sumar todos los diferentes productos de n elementos que se pueden formar con los elementos de A.
- El determinante ayuda a encontrar la inversa de una matriz

¿Cómo se calcula el determinante?

Para calcular el determinante, se debe tener una matriz cuadrada *nxn*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} |A| = ad - bc$$

Ejemplo:

$$B = \begin{bmatrix} 4 & 6 \\ 3 & 8 \end{bmatrix} |B| = 4 * 8 - 6 * 3 = 14$$

$|B| = 14$

Instanciar las variables para calcular el determinante

Calcular el determinante de:

$$|M - (\lambda I)| = 0$$

Se instancian todas las variables:

$$\left| \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

Se instancian todas las variables:

$$\left| \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

Se obtiene:

$$\left| \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$A - B = a_{i,j} - b_{i,j}$

$$\left| \begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix} \right| = 0$$

Calcular el determinante de:

$$\left| \begin{bmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix} \right| = 0$$

$|M - (\lambda I)| = (3-\lambda)(6-\lambda) - 4 = 0$
$|M - (\lambda I)| = 18 - 3\lambda - 6\lambda + \lambda^2 - 4 = 0$
$|M - (\lambda I)| = \lambda^2 - 9\lambda + 14 = 0$ *ecuación cuadrática*

Para resolver una ecuación cuadrática, usamos la función cuadrática:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Instanciar la función cuadrática

Para resolver una ecuación cuadrática, usamos la función cuadrática:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Instanciamos la ecuación cuadrática:

$$\lambda^2 - 9\lambda + 14 = 0$$

$$\lambda = \frac{-(-9) \pm \sqrt{9^2 - 4(1 \cdot 14)}}{2 \cdot 1}$$

Instanciamos la ecuación cuadrática:

$$\lambda = \frac{-(-9) \pm \sqrt{9^2 - 4(1 \cdot 14)}}{2 \cdot 1}$$

$$\lambda = \frac{9 \pm \sqrt{81 - 56}}{2}$$

$$\lambda = \frac{9 \pm \sqrt{25}}{2}$$

$$\lambda = \frac{9 \pm 5}{2}$$

Calcular el valor propio usando la función cuadrática

$$\lambda_1 = \frac{9+5}{2} = 7$$

$$\lambda_2 = \frac{9-5}{2} = 2$$

Los valores propios de la matriz M son: 7 y 2

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

Calcular el vector propio para el valor propio de $\lambda = 7$

Recordamos la ecuación a demostrar:

$$Mv = \lambda v$$

Sustituimos los valores que ya conocemos:

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix}$$

Sustituimos los valores que ya conocemos:

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix}$$

Del lado izquierdo, se multiplica la matriz con el vector y del lado derecho se multiplica el escalar con el vector.

$$3x + 2y = 7x$$
$$2x + 6y = 7y$$

$$3x + 2y = 7x$$
$$2x + 6y = 7y$$

Pasamos todo al lado izquierdo:

$$3x + 2y - 7x = 0$$
$$2x + 6y - 7y = 0$$
$$-4x + 2y = 0$$
$$2x - y = 0$$
$$y = 2x$$

Un posible vector propio es:

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

Dado que el posible vector, no es un vector unitario, debido a que la suma de los cuadrados de sus componentes es 5, en lugar de 1.

$$1^2 + 2^2 = 5$$

Por lo tanto:

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Un vector unitario es aquel donde la suma de cuadrados de sus componentes es 1.

$$(1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1$$

Por lo tanto, el vector propio es

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Y el valor propio para este vector es: 7

Usando los resultado obtenidos, demostrar que $Mv = \lambda v$

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Del lado izquierdo, multiplicamos M por v y del lado derecho se multiplica λ por v

$$\begin{bmatrix} 3/\sqrt{5} + 4/\sqrt{5} \\ 2/\sqrt{5} + 12/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

Finalmente, sumamos

$$\begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 7/\sqrt{5} \\ 14/\sqrt{5} \end{bmatrix}$$

- Ventajas: es posible calcular con exactitud los valores y vectores propios
- Desventaja: el tiempo de ejecución para calcular todos los pares de valores y vectores propios es $O(n^3)$

Método de las potencias: generalidades

- Es un método iterativo que aproxima los valores y vectores propios de una matriz.
- Este método se emplea en matrices grandes, particularmente Google emplea este método para calcular el PageRank de las páginas.
- Forma parte de los métodos libres de matrices.

Sea M una matriz, iniciamos con un vector v^0 e iteramos:

$$v_{k+1} := \frac{Mv_k}{||Mv_k||}$$

- Este proceso se hace iterativo hasta que converja $(||v_k - v_{k+1}||) < \epsilon$.
- El valor resultante de v es aproximadamente el vector propio de M.
- Para calcular el valor propio simplemente se aplica $\lambda_1 = v^T M v$ la cual se desprende de la ecuación de $Mv = \lambda v$, donde v es un vector unitario.

Las suposiciones importantes para que el método converja son:

- Existe un valor propio dominante es decir: $|v_1| > |v_i|$ para i = 1,2,...,n
- El vector inicial v^0 no puede ser ortogonal a v_1

- Dada la matriz cuadrada M, calcular los valores propios y los vectores propios usando el método de las potencias

$$M = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

Inicializamos el vector ($v0$) como un vector de 1s en ambos componentes:

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

Aplicando la norma de Frobenius el resultado es $\sqrt{5^2 + 8^2}$ = 9.434, se obtiene v_1:

$$v_1 = \begin{bmatrix} 5/9.434 \\ 8/9.434 \end{bmatrix} = \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix}$$

En la siguiente iteración obtenemos:

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.530 \\ 0.848 \end{bmatrix} = \begin{bmatrix} 3.286 \\ 6.148 \end{bmatrix}$$

Aplicando nuevamente la norma de Frobenius el resultado es $\sqrt{3.286^2 + 6.148^2}$ = 6.971,

$$v_2 = \begin{bmatrix} 0.471 \\ 0.882 \end{bmatrix}$$

Después de k iteraciones, obtenemos el siguiente vector cuyo segundo elemento es dos veces el primero:

$$v = \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix}$$

Este es el vector principal de la matriz M.

Para obtener el valor principal dado el vector previo, únicamente se calcula:

$$\lambda = v^T M v = \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} = 6.993$$

Para encontrar el segundo par (vector y valor propio), se crea una nueva matriz $M^* = M - \lambda v v^T$

$$M^* = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - 6.993 \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \end{bmatrix}$$

$$M^* = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} 1.397 & 2.795 \\ 2.795 & 5.589 \end{bmatrix} = \begin{bmatrix} 1.603 & -0.795 \\ -0.795 & 0.411 \end{bmatrix}$$

Para calcular el segundo par de vectores y valores propios es necesario procesar la matriz M^* como se hizo con la matriz original M.

Si M^* = $M - \lambda vv^T$ donde v y λ son un par de vector y valor propio con el valor propio más grande, entonces:

- v es también un vector propio de M^* y su correspondiente valor propio es 0. Dado que:

$$M^*v = (M - \lambda vv^T)v = Mv - \lambda vv^Tv = Mv - \lambda v = 0$$

En el penúltimo paso usamos el hecho que $v^Tv = 1$ debido a que v es un vector unitario

- En cambio, si v_1 y $\lambda_v 1$ son un par de valor y vector propio de una matriz simétrica distinta del primer par propio (v, λ), entonces también son un par de M^*, la prueba es:

$$M^*v = (M^*)^T v = (M - \lambda x x^T)^T v = M^T v - \lambda x(x^T v) = M^T v = \lambda_u v$$

Esta secuencia de igualdades sucede debido a:

- Si M es simétrica, entonces $M = M^T$
- Los vectores propios de una matriz simétrica son ortogonales. Esto es, el producto punto de dos vectores propios distintos de una matriz es 0

Algunos inconvenientes de este método son:

- En cada iteración es necesario vigilar que v^k sea distinta de cero, dado que en la siguiente iteración se utiliza este valor en la división.
- Los componentes de los vectores pueden crecer de forma desproporcionada o bien tender a cero

Peña, Método de potencia directo e indirecto, 2011.
https://www.cimat.mx/~joaquin/mn11/clase12.pdf