Datos Masivos II

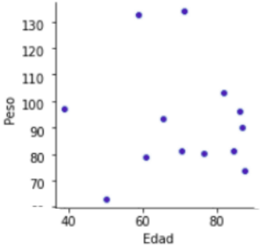
Reducción de dimensionalidad: Vectores y valores propios de matrices simétricas

Blanca Vázquez Gómez

Agosto 2020

Introducción

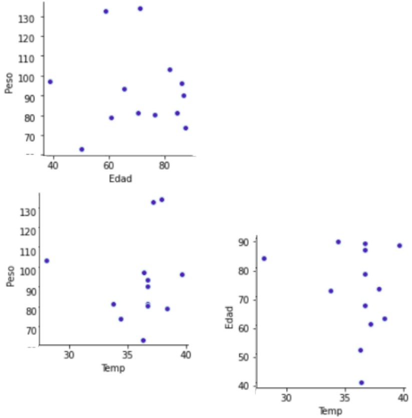
Paciente	Edad	Peso	Temp	Colesterol	Sodio	Glucosa	Ritmo cardiaco	…
1	88.78	74.00	34.40	96.00	132.33	137.35	73.77	93.19
2	72.33	134.00	37.90	168.35	132.50	124.00	85.25	81.18
3	88.27	90.00	36.71	188.00	134.00	137.35	75.85	81.18
4	66.68	93.19	36.71	116.00	139.50	137.35	58.82	103.00
5	71.70	81.18	33.77	168.35	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

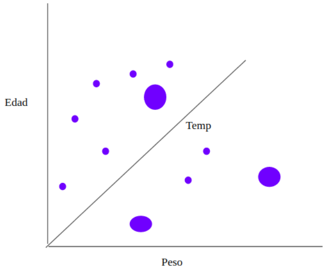
Una base de datos, con k número de atributos, es posible visualizar los datos como una nube de puntos en un espacio de k-dimensiones.

Introducción



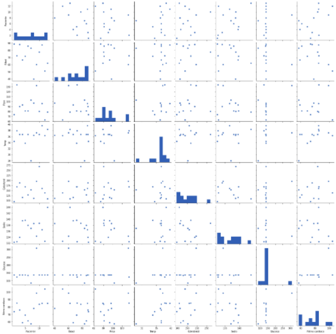
Paciente	Edad	Peso	Temp
1	88.78	74.00	34.40
2	72.33	134.00	37.90
3	88.27	90.00	36.71
4	66.68	93.19	36.71
5	71.70	81.18	33.77
6	85.92	81.18	36.71
7	83.09	103.00	28.03
8	39.91	97.00	36.40
9	60.15	132.50	37.20
10	77.69	80.20	36.71
11	87.45	96.20	39.60
12	51.26	63.10	36.30
13	62.16	78.90	38.38

Introducción



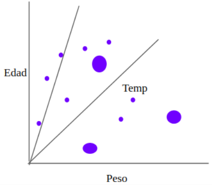
Paciente	Edad	Peso	Temp
1	88.78	74.00	34.40
2	72.33	134.00	37.90
3	88.27	90.00	36.71
4	66.68	93.19	36.71
5	71.70	81.18	33.77
6	85.92	81.18	36.71
7	83.09	103.00	28.03
8	39.91	97.00	36.40
9	60.15	132.50	37.20
10	77.69	80.20	36.71
11	87.45	96.20	39.60
12	51.26	63.10	36.30
13	62.16	78.90	38.38

Introducción



¿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

Técnicas para reducción de dimensionalidad

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)

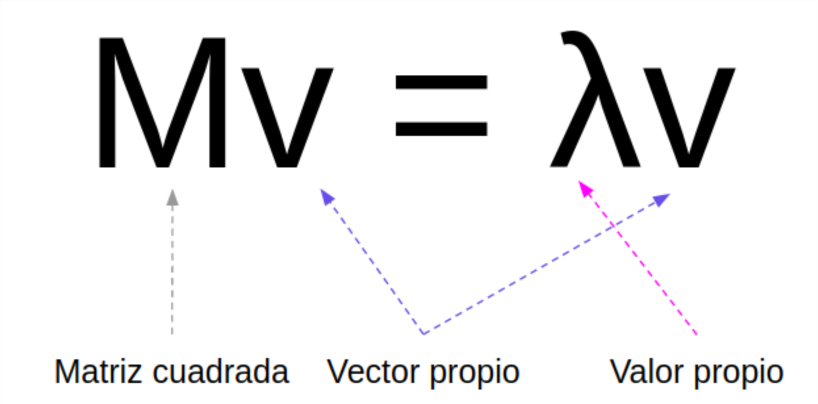
- Es una de las técnicas más comunes de reducción de dimensionalidad.
- Es un método de proyección dónde las m columnas de una base de datos se proyectan a un sub-espacio con m o menor número de características, conservando la esencia de los datos originales.
- Se calcula usando una matriz de descomposición, como la de descomposición propia.

Descomposición propia

Es la descomposición de una matriz cuadrada en un conjunto de vectores propios (eigenvectores) y valores propios (eigenvalores).

Definición

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


Mv = λv
Matriz cuadrada
Vector propio
Valor propio

Ejercicio

- 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 $Mv = \lambda v$

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

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.

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: $(\lambda I)v$

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

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:

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

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.

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

Calcular los vectores propios

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

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

Demostrar que $Mv = \lambda v$

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

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

Ejercicio

¿Cuál es el vector propio de la matriz M, cuando su valor propio es 2?

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