

Cuestionario PCA

Estadística Multivariada

Yair Castillo, Emilio Valencia, Alec Torres

Abril 2024

1 Ejercicio 1

1. Verdadero: Todas las componentes principales son ortogonales entre ellas, y por estar en la cámara de Weyl maximiza la variabilidad restante.
2. Verdadero: Por definición son no correlacionados, ya que son ortogonales.
3. Verdadero: Por construcción la dimensión de los datos transformados siempre es menor a la dimensión de los datos originales

2 Pasos para elaborar un PCA

1. Tomar la matriz y restarle la media por columnas.
2. Multiplicar la nueva matriz por su transpuesta.
3. Multiplicar esta matriz cuadrada y simétrica por $\frac{1}{n-1}$.
4. Calcular eigenvalores y eigenvectores.
5. Ordenarlos en la cámara de Weyl.

3 Ejercicio 2

1. Es una matriz diagonal, por lo que sus eigenvalores son: $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 2$.
2. Resolver el sistema:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \mathbf{v} = \lambda_1 \mathbf{v}$$

3. La primer componente principal es:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

4 Ejercicio 3

Observación	X1	X2
1	-2	2
2	2	-2

1. Sabemos que $b_{1,1}^2 + b_{1,2}^2 = 1$
2. Despejando podemos obtener que $b_{1,2} = \sqrt{1 - b_{1,1}^2}$
3. Por lo que: $Xb_1 = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 0.7071 \\ -0.5412 \end{pmatrix} = \begin{pmatrix} -2.4966 \\ 2.4966 \end{pmatrix}$

5 Ejercicio 4

Esta figura nos presenta las coordenadas de los datos originales y los compara con las coordenadas de los datos transformados, en las que se pueden observar la ortogonalidad de las componentes principales, así como las proyecciones generadas por cada dato.

6 Ejercicio 5

1. Vemos que, $Y = (Y_1, Y_2)^T$ se puede expresar como:

$$Y = \begin{pmatrix} 0.5 & 2 \\ -0.5 & 2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

La media de Y es:

$$E[Y] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

La matriz de covarianza de Y es:

$$\text{Cov}(Y) = \begin{pmatrix} 4.25 & 3.75 \\ 3.75 & 4.25 \end{pmatrix}$$

Por lo tanto, la distribución de Y es una distribución gaussiana multivariante con media $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ y matriz de covarianza $\begin{pmatrix} 4.25 & 3.75 \\ 3.75 & 4.25 \end{pmatrix}$. Los eigenvalores de la matriz de covarianza de Y son $\lambda_1 = 8$ y $\lambda_2 = \frac{1}{2}$.

7 Ejercicio 6

La varianza total se calcula como:

$$\text{Varianza total} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

donde:

X_i : Valores de la muestra

\bar{X} : Media de la muestra

n : Tamaño de la muestra

Dado que los datos están estandarizados y la media es cero, la varianza total se calcula como:

$$\text{Varianza total} = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Para la primera variable X_1 , utilizando los datos proporcionados:

$$(-0.577)^2 + (-0.577)^2 + (-0.577)^2 + (1.732)^2 = 0.333 + 0.333 + 0.333 + 3 = 4$$

Para la segunda variable X_2 :

$$(1)^2 + (1)^2 + (-1)^2 + (-1)^2 = 1 + 1 + 1 + 1 = 4$$

Para la tercera variable X_3 :

$$(-1)^2 + (1)^2 + (1)^2 + (-1)^2 = 1 + 1 + 1 + 1 = 4$$

Sumamos estos valores para obtener el numerador correcto:

$$4 + 4 + 4 = 12$$

Ahora, volvemos a calcular la varianza total:

$$\text{Varianza Total} = \frac{12}{12} = 1$$

Luego, calculamos el autovalor asociado con el autovector dado:

$$\text{Autovalor} = (0.707^2 + (-0.5)^2 + (-0.5)^2) \times 1 = 0.999$$

Ahora, calculamos la proporción de la varianza explicada por la primera componente principal:

$$\text{PVE} = \frac{0.999}{1} = 0.999$$

Por lo tanto, la proporción de la varianza explicada por la primera componente principal es aproximadamente del 99.9%.

8 Ejercicio 7

1. Verdadero, dado que las componentes están en la cámara de Weyl, cada una explica menos varianza que la anterior.
2. Verdadero, por lo mismo que el inciso 1.
3. Verdadero, cada componente adicional explica más los datos.
4. Verdadero, esta gráfica nos permite ver cuantas componentes principales se deben usar para explicar la mayor parte de la varianza sin perder mucha información.

9 Ejercicio 8

Las 3 aservaciones pueden ser demostradas visualmente.

1. Para la primera basta observar que el ángulo en la gráfica escalada entre los vectores X_1 y X_2 es mucho menor al ángulo formado por los vectores X_1 y X_3 , lo que indica que X_1 está más correlacionado con X_2 que con X_3 .
2. Para la segunda notemos que en la gráfica escalada el vector X_3 es el de mayor longitud, quizás casi hasta el doble de largo que X_1, X_2 y X_4 , por ello sabemos que la varianza más alta es la de X_3 .
3. Y para la tercera observación notemos en la gráfica escalada que la observación 24 está en el extremo positivo de X_4 , es decir, dicha observación es relativamente grande par X_4 y es positiva.

10 Ejercicio 9

1. Es válido si los dos primeros componentes principales recogen la misma cantidad total de variabilidad en los datos. Esto se debe a que los componentes PC_1 y PC_2 son perpendiculares entre sí y representan diferentes direcciones de variabilidad. Por lo tanto, la suma total de la variabilidad recogida por ambos componentes principales es igual a la varianza total de los datos.

2. Esta declaración se centra en la covarianza entre los coeficientes asociados a los dos primeros componentes principales. Si ambos componentes son perpendiculares entre sí, su covarianza es nula. Dado que tenemos la certeza de que ξ_1 y ξ_2 son perpendiculares, la afirmación es correcta.

3. Se está discutiendo la suma de los cuadrados de los coeficientes de cada componente principal. Cuando los componentes están estandarizados, la suma de los cuadrados de los coeficientes de cada componente principal es igual a 1, lo que resulta en una varianza de cada componente principal igual a 1. Por lo tanto, la afirmación es verdadera.

11 Ejercicio 10

La opción D es la mejor ya que como sabemos la primera componente principal (es la máxima debido a la cámara de Weigl) representa la máxima variabilidad de los datos. Las demás opciones no pueden ser ya que tienen todos los elementos iguales y están distribuidos de manera similar, o tiene elementos alternados positivos y negativos.