Métodos Computacionales

Clase IX: Optimización Restringida y Análisis de Componentes Principales

Optimización Restringida

Optimización restringida

- El problema consiste en encontrar el valor máximo (o mínimo) de una forma cuadrática $Q(\mathbf{x})$ para un $\mathbf{x}$ en un conjunto específico.

- Sin pérdida de generalidad el problema se puede plantear para todos los $\mathbf{x}$ que sean vectores unitarios, es decir $||\mathbf{x}|| = 1$.

Optimización restringida

- El problema consiste en encontrar el valor máximo (o mínimo) de una forma cuadrática $Q(\mathbf{x})$ para un $\mathbf{x}$ en un conjunto específico.

- Sin pérdida de generalidad el problema se puede plantear para todos los $\mathbf{x}$ que sean vectores unitarios, es decir $||\mathbf{x}|| = 1$.

$$\|\mathbf{x}\| = 1, \quad \|\mathbf{x}\|^2 = 1, \quad \mathbf{x}^T\mathbf{x} = 1$$
$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$$

Optimización restringida: ejemplo

- Obtener el máximo y mínimo de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

Forma cuadrática (sin productos cruzados):

$$A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Optimización restringida: ejemplo

- Obtener el **máximo** y mínimo de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{array}{l} 4x_2^2 \leq 9x_2^2 \\ 3x_3^2 \leq 9x_3^2 \end{array}\right\} \quad Q(\mathbf{x}) \leq 9x_1^2 + 9x_2^2 + 9x_3^2 \longrightarrow Q(\mathbf{x}) \leq 9(x_1^2 + x_2^2 + x_3^2)$$

Optimización restringida: ejemplo

- Obtener el **máximo** y mínimo de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$||\mathbf{x}|| = x_1^2 + x_2^2 + x_3^2 = 1$$

$$\left.\begin{array}{l} 4x_2^2 \leq 9x_2^2 \\ 3x_3^2 \leq 9x_3^2 \end{array}\right\} \quad Q(\mathbf{x}) \leq 9x_1^2 + 9x_2^2 + 9x_3^2 \longrightarrow Q(\mathbf{x}) \leq 9\underbrace{(x_1^2 + x_2^2 + x_3^2)}$$

Optimización restringida: ejemplo

- Obtener el **máximo** y mínimo de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{aligned} 4x_2^2 &\leq 9x_2^2 \\ 3x_3^2 &\leq 9x_3^2 \end{aligned}\right\} \quad Q(\mathbf{x}) \leq 9x_1^2 + 9x_2^2 + 9x_3^2 \longrightarrow Q(\mathbf{x}) \leq 9$$

Optimización restringida: ejemplo

- Obtener el **máximo** y mínimo de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{array}{l} 4x_2^2 \leq 9x_2^2 \\ 3x_3^2 \leq 9x_3^2 \end{array}\right\} \; Q(\mathbf{x}) \leq 9x_1^2 + 9x_2^2 + 9x_3^2 \longrightarrow Q(\mathbf{x}) \leq 9$$

En particular: $x = (1, 0, 0) \to Q(1, 0, 0) = 9$

Optimización restringida: ejemplo

- Obtener el máximo y **mínimo** de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{aligned} 4x_2^2 &\geq 3x_2^2 \\ 9x_1^2 &\geq 3x_1^2 \end{aligned}\right\} \quad Q(\mathbf{x}) \geq 3x_1^2 + 3x_2^2 + 3x_3^2 \longrightarrow Q(\mathbf{x}) \geq 3(x_1^2 + x_2^2 + x_3^2)$$

Optimización restringida: ejemplo

- Obtener el máximo y **mínimo** de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{array}{l} 4x_2^2 \geq 3x_2^2 \\ 9x_1^2 \geq 3x_1^2 \end{array}\right\} \quad Q(\mathbf{x}) \geq 3x_1^2 + 3x_2^2 + 3x_3^2 \longrightarrow Q(\mathbf{x}) \geq 3(\overbrace{x_1^2 + x_2^2 + x_3^2}^{||\mathbf{x}|| = x_1^2 + x_2^2 + x_3^2 = 1})$$

Optimización restringida: ejemplo

- Obtener el máximo y **mínimo** de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{aligned} 4x_2^2 &\geq 3x_2^2 \\ 9x_1^2 &\geq 3x_1^2 \end{aligned}\right\} \quad Q(\mathbf{x}) \geq 3x_1^2 + 3x_2^2 + 3x_3^2 \longrightarrow Q(\mathbf{x}) \geq 3$$

Optimización restringida: ejemplo

- Obtener el máximo y **mínimo** de Q (sujeto a $||\mathbf{x}|| = 1$):

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\left.\begin{aligned} 4x_2^2 &\geq 3x_2^2 \\ 9x_1^2 &\geq 3x_1^2 \end{aligned}\right\} \quad Q(\mathbf{x}) \geq 3x_1^2 + 3x_2^2 + 3x_3^2 \longrightarrow Q(\mathbf{x}) \geq 3$$

En particular: $x = (0, 0, 1) \rightarrow Q(0, 0, 1) = 3$

Optimización restringida

- Si la ecuación cuadrática no tiene términos cruzados es fácil de mostrar que:
  - El **máximo** (restringido) de $Q(\mathbf{x})$ es igual al **mayor autovalor**
  - El **mínimo** (restringido) de $Q(\mathbf{x})$ es igual al **menor autovalor**

Optimización restringida

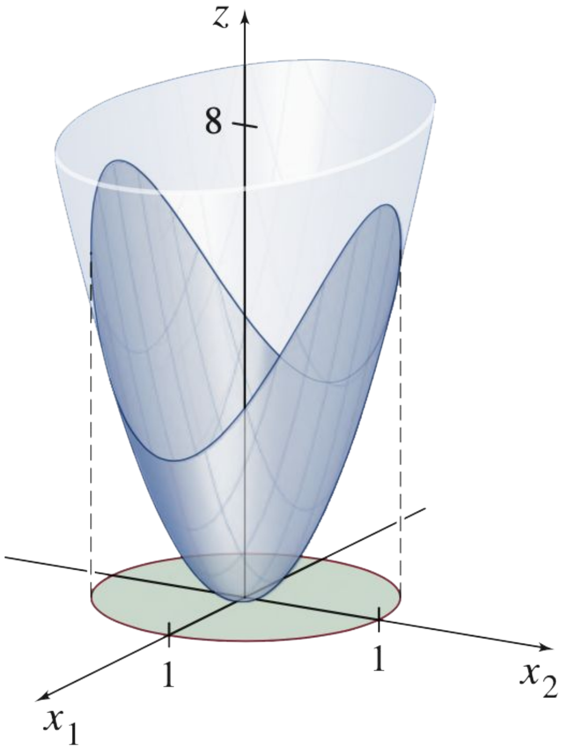
- Si la ecuación cuadrática no tiene términos cruzados es fácil de mostrar que:
  - El **máximo** (restringido) de $Q(\mathbf{x})$ es igual al **mayor autovalor**
  - El **mínimo** (restringido) de $Q(\mathbf{x})$ es igual al **menor autovalor**
- Veremos que esto también es válido para cualquier forma cuadrática $Q(\mathbf{x})$

Intuición visual

- Forma cuadrática:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$$

$$z = 3x_1^2 + 7x_2^2$$



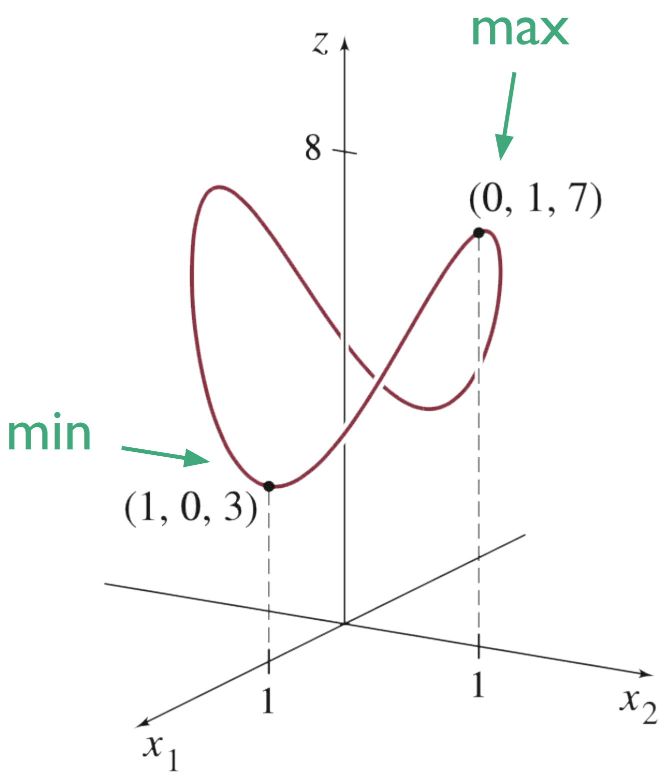
Intuición visual

- Forma cuadrática:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$$

$$z = 3x_1^2 + 7x_2^2$$

$$\text{sujeto a } x_1^2 + x_2^2 = 1$$



Optimización restringida

- Sea A una matriz simétrica, definimos:

$$m = \min \left\{ \mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1 \right\}$$
$$M = \max \left\{ \mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1 \right\}$$

Optimización restringida

- Sea A una matriz simétrica, definimos:

$$m = \min\left\{\mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1\right\}$$
$$M = \max\left\{\mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1\right\}$$

Entonces:

$m = \lambda_n$ (el valor propio más chico de A)

$M = \lambda_1$ (el valor propio más grande de A)

Optimización restringida: teorema

- Sea A una matriz simétrica:
  - El **máximo restringido** de la forma cuadrática es el mayor autovalor de la matriz A (λ_1) y este máximo se cumple en el vector propio unitario $\mathbf{u}_1$
  - El **mínimo restringido** de la forma cuadrática es el menor autovalor de la matriz A (λ_n) y este mínimo se cumple en el vector propio unitario $\mathbf{u}_n$

Optimización restringida

- Sea A una matriz simétrica, definimos:

$$m = \min\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\}$$
$$M = \max\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\}$$

Entonces:

$$m = \lambda_n \qquad \underline{\mathbf{u}_n^T A\mathbf{u}_n = \lambda_n}$$
$$M = \lambda_1 \qquad \underline{\mathbf{u}_1^T A\mathbf{u}_1 = \lambda_1}$$

Demostración intuitiva

- Sabiendo que:

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} \quad \text{cuando } \mathbf{x} = P\mathbf{y}$$

y además:

$$\|\mathbf{x}\| = \|P\mathbf{y}\| = \|\mathbf{y}\|$$

entonces $\mathbf{x}^T A\mathbf{x}$ y $\mathbf{y}^T D\mathbf{y}$ toman el mismo conjunto de valores cuando $\mathbf{x}$ e $\mathbf{y}$ recorren los vectores unitarios.

Con este cambio de variable y la demostración para la matriz diagonal, podemos **extender** la demostración para **toda matriz simétrica** A.

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^\mathbf{T} A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^\mathbf{T}\mathbf{x} = 1$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

$$\begin{aligned} 0 &= -\lambda^3 + 10\lambda^2 - 27\lambda + 18 \\ &= -\underbrace{(\lambda - 6)}_{\text{Máximo autovalor, } \lambda = 6}(\lambda - 3)(\lambda - 1) \end{aligned}$$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}} A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

$$\begin{aligned} 0 &= -\lambda^3 + 10\lambda^2 - 27\lambda + 18 \\ &= -\underbrace{(\lambda - 6)}_{\text{Máximo autovalor, } \lambda = 6}(\lambda - 3)(\lambda - 1) \end{aligned}$$

Resolviendo $(A - 6\, I)\,\mathbf{x} = 0 \longrightarrow \mathbf{x} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\text{Autovector asociado a } \lambda = 6}$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$

- El máximo restringido de $\mathbf{x}^{\mathrm{T}}A\mathbf{x}$ se obtiene cuando $\mathbf{x}$ es un vector propio **unitario** para $\lambda = 6$:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \longrightarrow \mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Teorema (continuación)

- Sea A una matriz simétrica, λ_1 su mayor autovalor y $\mathbf{u_1}$ el vector propio asociado, entonces **el valor máximo** de $\mathbf{x^T}A\mathbf{x}$ sujeto a las restricciones:

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

es el segundo valor propio más grande λ_2 y se alcanza este máximo cuando $\mathbf{x}$ es el vector propio $\mathbf{u_2}$ correspondiente a λ_2

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$ y $\mathbf{x}^{\mathbf{T}}\mathbf{u}_{\mathbf{1}} = 0$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix} \quad \mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$ y $\mathbf{x}^{\mathbf{T}}\mathbf{u_1} = 0$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

$$\begin{aligned} 0 &= -\lambda^3 + 10\lambda^2 - 27\lambda + 18 \\ &= -(\lambda - 6)\underbrace{(\lambda - 3)}_{\text{Segundo autovalor más grande } \lambda = 3}(\lambda - 1) \end{aligned}$$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$ y $\mathbf{x}^{\mathbf{T}}\mathbf{u}_{\mathbf{1}} = 0$

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

$$\begin{aligned} 0 &= -\lambda^3 + 10\lambda^2 - 27\lambda + 18 \\ &= -(\lambda - 6)\underbrace{(\lambda - 3)}_{\text{Segundo autovalor más grande } \lambda = 3}(\lambda - 1) \end{aligned}$$

Resolviendo $(A - 3I)\,\mathbf{x} = 0 \longrightarrow \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$

Ejemplo

- Encontrar el máximo de la forma cuadrática $\mathbf{x}^{\mathbf{T}} A\mathbf{x}$ sujeto a la restricción $\mathbf{x}^{\mathbf{T}}\mathbf{x} = 1$ y $\mathbf{x}^{\mathbf{T}}\mathbf{u}_{\mathbf{1}} = 0$
- El máximo es el vector propio **unitario** asociado al segundo mayor autovalor $\lambda = 3$:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \longrightarrow \mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

Teorema (generalización)

- Sea A una matriz simétrica de n x n con una diagonalización ortogonal $A=PDP^{-1}$ donde las entradas sobre la diagonal D están acomodadas de manera que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ y donde las columnas de P son vectores propios unitarios correspondientes $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Entonces para $k = 2, \ldots, n$, el valor máximo de $\mathbf{x}^T A\mathbf{x}$ sujeto a las restricciones:

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0, \quad \ldots, \quad \mathbf{x}^T\mathbf{u}_{k-1} = 0$$

es el valor propio λ_k, y alcanza su máximo en $\mathbf{x} = \mathbf{u}_k$

Descomposición en Valores Singulares

Descomposición en Valores Singulares

- Nos gustaría poder descomponer cualquier matriz como

$$A = PDP^{-1}$$

- Podemos encontrar una **Descomposición en Valores Singulares** $A=QDP^{-1}$ para cualquier matriz de m x n.

Propiedad

- Los valores absolutos de los valores propios de una matriz simétrica A mide la cantidad de "estiramiento" que produce A sobre los vectores propios.

Propiedad

- Los valores absolutos de los valores propios de una matriz simétrica A mide la cantidad de "estiramiento" que produce A sobre los vectores propios.

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\| = |\lambda|$$

Propiedad

- Los valores absolutos de los valores propios de una matriz simétrica A mide la cantidad de "estiramiento" que produce A sobre los vectores propios.

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\| = |\lambda|$$

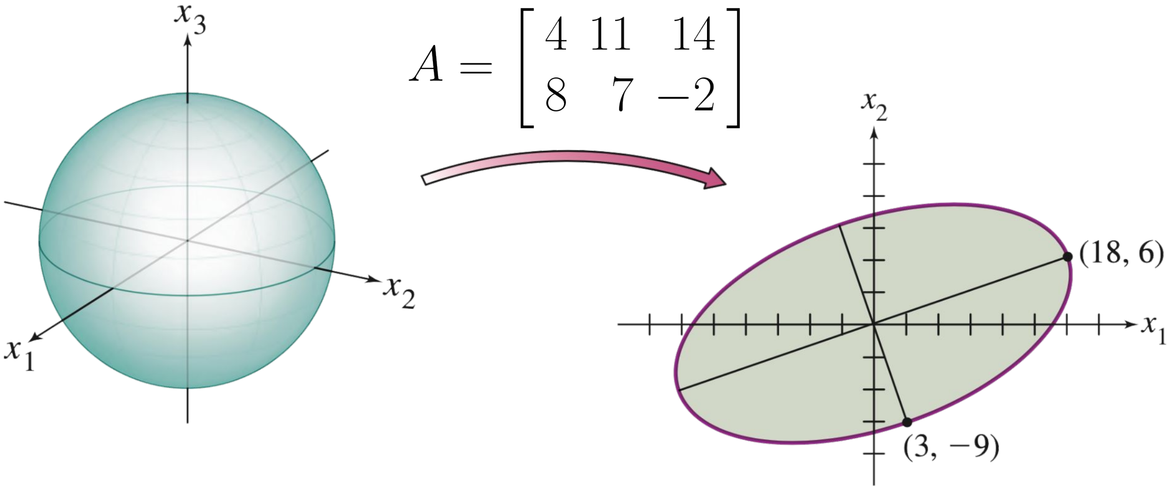
- Si λ_1 es el valor propio de mayor magnitud, su vector propio asociado $\mathbf{v}_1$ es la dirección en la que el "estiramiento" es máximo.

Descomposición en Valores Singulares

- Es posible imitar esta propiedad para matrices rectangulares.

Ejemplo

- Ejemplo: encontrar un vector unitario $\mathbf{x}$ que maximice la longitud $||A\mathbf{x}||$, ¿Cuál es esa longitud?



Ejemplo

- La cantidad $||A\mathbf{x}||$ se maximiza con la misma $\mathbf{x}$ que maximiza $||A\mathbf{x}||^2$, entonces:

Ejemplo

- La cantidad $||A\mathbf{x}||$ se maximiza con la misma $\mathbf{x}$ que maximiza $||A\mathbf{x}||^2$, entonces:

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \left(A^T A\right) \mathbf{x}$$

Ejemplo

- La cantidad $||A\mathbf{x}||$ se maximiza con la misma $\mathbf{x}$ que maximiza $||A\mathbf{x}||^2$, entonces:

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \underbrace{\mathbf{x}^T \underbrace{\left(A^T A\right)}_{\text{Matriz simétrica}} \mathbf{x}}_{\text{Forma cuadrática}}$$

Ejemplo

- Calculamos $A^{\mathrm{T}}A$:

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Ejemplo

- Calculamos $A^{\mathrm{T}}A$, sus valores y vectores propios:

$$\lambda_1 = 360, \lambda_2 = 90, \ \text{and } \lambda_3 = 0$$

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

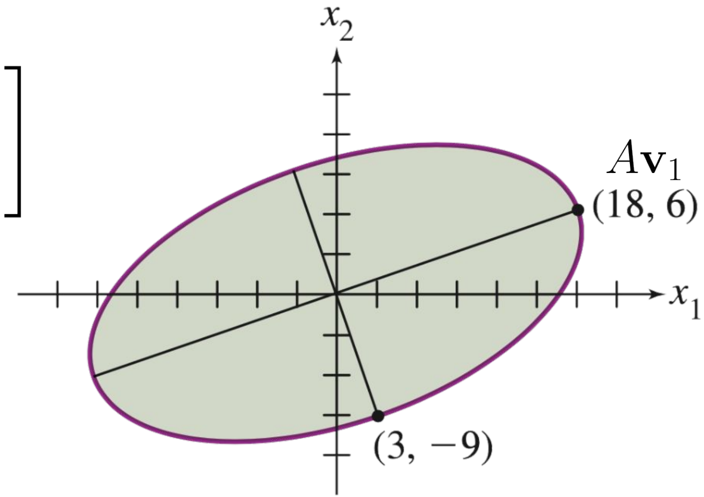
Ejemplo

- El valor máximo de $||A\mathbf{x}||^2$ es 360, que se alcanza cuando $\mathbf{x}$ es el vector unitario $\mathbf{v}_1$.

Ejemplo

- El valor máximo de $||A\mathbf{x}||^2$ es 360, que se alcanza cuando $\mathbf{x}$ es el vector unitario $\mathbf{v}_1$.
- El vector $A\mathbf{v}_1$ es el punto sobre la elipse más alejado del origen:

$$A\mathbf{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}$$

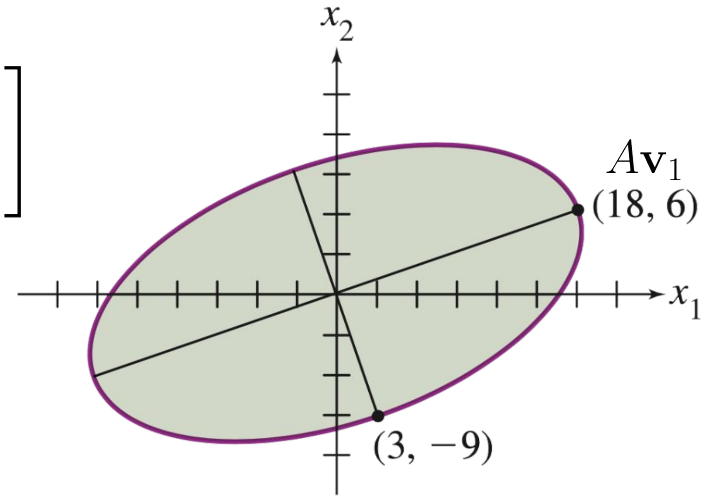


Ejemplo

- El valor máximo de $||A\mathbf{x}||^2$ es 360, que se alcanza cuando $\mathbf{x}$ es el vector unitario $\mathbf{v}_1$.
- El vector $A\mathbf{v}_1$ es el punto sobre la elipse más alejado del origen:

$$A\mathbf{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}$$

$$||A\mathbf{v}_1|| = \underbrace{\sqrt{360} = 6\sqrt{10}}_{\text{Valor máximo (sujeto a } ||\mathbf{x}||=1)}$$



Valores singulares

- Los **valores singulares** de una matriz A de m x n son las raíces cuadradas de los valores propios de $A^{\mathrm{T}}A$, que denotamos en orden decreciente:

$$\sigma_1, \ldots, \sigma_n$$

Es decir:

$$\sigma_i = \sqrt{\lambda_i}, \ \text{para } 1 \leq i \leq n$$

Valores singulares: ejemplo

- Para nuestra matriz A del ejemplo anterior:

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

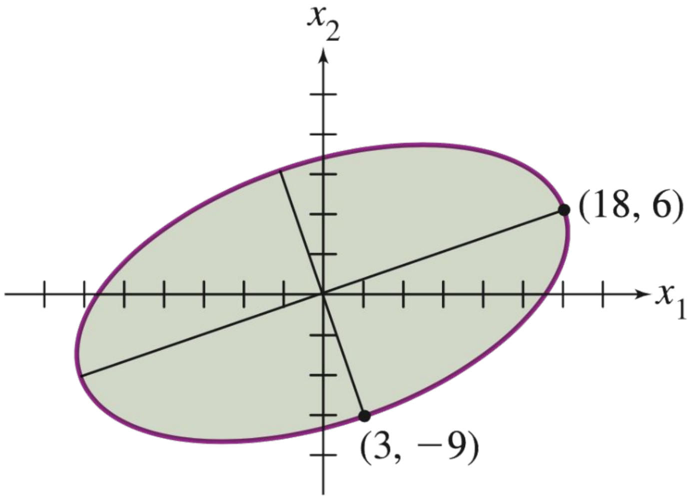
$$\sigma_1 = \sqrt{360} = 6\sqrt{10}$$

$$\sigma_2 = \sqrt{90} = 3\sqrt{10}$$

$$\sigma_3 = 0$$

Valores singulares: ejemplo

- El primer valor singular de A es el máximo de $||A\mathbf{x}||$ sobre todos los vectores unitarios, y se alcanza en el vector propio unitario $\mathbf{v}_1$.
- El segundo valor singular de A es el máximo de $||A\mathbf{x}||$ sobre todos los vectores unitarios **ortogonales** a $\mathbf{v}_1$, y se alcanza en el segundo vector propio unitario $\mathbf{v}_2$.



$$\sigma_1 = \sqrt{360} = 6\sqrt{10}$$
$$\sigma_2 = \sqrt{90} = 3\sqrt{10}$$

Valores singulares

- Si $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ es una base ortonormal de $\mathbb{R}^n$ que consiste en los vectores propios de $A^T A$, arreglados de tal forma que los valores propios satisfacen $\lambda_1 \geq \ldots \geq \lambda_n$, y A tiene r valores singulares diferentes de cero, entonces $\{A\mathbf{v}_1, ..., A\mathbf{v}_r\}$ es una base ortogonal para $\mathrm{Col}\ A$, y rango $A = r$.

Descomposición en Valores Singulares (SVD)

- Dada una matriz A de m x n con rango r, entonces existe una matriz Σ de m x n, para la cual las entradas diagonales en D son los primeros r valores singulares de A, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, y existen una matriz ortogonal U de m x m y una matriz ortogonal V de n x n tales que:

$$A = U\Sigma V^T \qquad \Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \rightarrow m - r \text{ filas } 0$$

n - r columnas 0

Descomposición en Valores Singulares (SVD)

- Dada una matriz A de m x n con rango r, entonces existe una matriz Σ de m x n, para la cual las entradas diagonales en D son los primeros r valores singulares de A, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, y existen una matriz ortogonal U de m x m y una matriz ortogonal V de n x n tales que:

$$A = \underbrace{U}_{\text{vectores singulares izquierdos (columnas de } U)}\Sigma \underbrace{V^T}_{\text{vectores singulares derechos (columnas de } V)}$$

$$\Sigma = \begin{array}{cc} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} & \rightarrow m - r \text{ filas } 0 \\ \downarrow n - r \text{ columnas } 0 & \end{array}$$

vectores singulares izquierdos (columnas de U)

vectores singulares derechos (columnas de V)

Descomposición en Valores Singulares (SVD)

- Construir una descomposición en valores singulares de:

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

1. Diagonalizar ortogonalmente a $A^{\mathrm{T}}A$, encontrar sus valores y vectores propios:

$$\lambda_1 = 360,\ \lambda_2 = 90,\ \text{and}\ \lambda_3 = 0$$

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

2. Construir V acomodando los vectores propios de $A^{\mathrm{T}}A$ en orden decreciente según sus correspondientes valores propios:

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

3. A partir de los valores propios de $A^{\mathrm{T}}A$, calcular los valores singulares de A (raíces cuadradas de los valores propios de $A^{\mathrm{T}}A$) y formar las matrices D y Σ:

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix} \rightarrow \Sigma = \begin{bmatrix} D & 0 \end{bmatrix} = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1$, ..., $A\mathbf{v}_r$ (con r rango de A).

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1$, ..., $A\mathbf{v}_r$ (con r rango de A).

$$U = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

5. Finalmente:

$$A = \underbrace{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}_{V^T}$$

Descomposición en Valores Singulares (SVD)

- Encontrar una descomposición en valores singulares de:

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

Descomposición en Valores Singulares (SVD)

1. Diagonalizar ortogonalmente a $A^{\mathrm{T}}A$, encontrar sus valores y vectores propios:

$$A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix} \longrightarrow \lambda_1 = 18, \lambda_2 = 0$$

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

2. Construir V acomodando los vectores propios de $A^{\mathrm{T}}A$ en orden decreciente según sus correspondientes valores propios:

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

3. A partir de los valores propios de $A^{\mathrm{T}}A$, calcular los valores singulares de A (raíces cuadradas de los valores propios de $A^{\mathrm{T}}A$) y formar las matrices D y Σ:

$$\sigma_1 = 3\sqrt{2}, \quad \sigma_2 = 0$$

$$D = [3\sqrt{2}]$$

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1$, ..., $A\mathbf{v}_r$ (con r rango de A).

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \mathbf{u}_1 = \frac{1}{3\sqrt{2}} A\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1$, ..., $A\mathbf{v}_\mathrm{r}$ (con r rango de A). Ampliamos el conjunto $\{\mathbf{u}_1\}$ a una base ortonormal para $\mathbb{R}^3$:

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1$, ..., $A\mathbf{v}_r$ (con r rango de A).
   Ampliamos el conjunto $\{\mathbf{u}_1\}$ a una base ortonormal para $\mathbb{R}^3$:

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \perp \mathbf{u}_1$$

Descomposición en Valores Singulares (SVD)

4. Construir U: las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1, ..., A\mathbf{v}_r$ (con r rango de A).
   Ampliamos el conjunto $\{\mathbf{u}_1\}$ a una base ortonormal para $\mathbb{R}^3$:

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

Vectores $\mathbf{w}_1$ y $\mathbf{w}_2$ ortogonalizados (Gram-Schmidt) y normalizados

Descomposición en Valores Singulares (SVD)

5. Finalmente:

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V^T}$$

Aplicaciones

Aplicaciones al procesamiento de datos

- Una de las aplicaciones más extendidas de la descomposición SVD gira en torno a una técnica llamada **Análisis de Componentes Principales** (PCA).

Aplicaciones al procesamiento de datos

- Una de las aplicaciones más extendidas de la descomposición SVD gira en torno a una técnica llamada **Análisis de Componentes Principales** (PCA).
- Este análisis es aplicable a cualquier conjunto de datos que consista en una **lista de mediciones**.

Ejemplo PCA: conjunto de datos

- Supongamos un conjunto de datos con pesos y alturas de N estudiantes de la LTD.
- Llamamos $\mathbf{X}_j$ al vector de observaciones en $\mathbb{R}^2$ que lista el peso y la altura del j-ésimo estudiante:

$$\begin{array}{cccc} & & & \\ \left[\begin{array}{cccc} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \end{array}\right] \\ \begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ \mathbf{X}_1 & \mathbf{X}_2 & & \mathbf{X}_N \end{array} \end{array}$$

Ejemplo PCA: conjunto de datos

- Supongamos un conjunto de datos con pesos y alturas de N estudiantes de la LTD.
- Llamamos $\mathbf{X}_j$ al vector de observaciones en $\mathbb{R}^2$ que lista el peso y la altura del j-ésimo estudiante:

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \end{bmatrix}$$

Diagrama de dispersión

Ejemplo PCA: media y covarianza

- Como preparación al análisis de componentes principales, definimos la **media muestral $\mathbf{M}$** como:

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

Ejemplo PCA: media y covarianza

- Como preparación al análisis de componentes principales, definimos la **media muestral M** como:

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

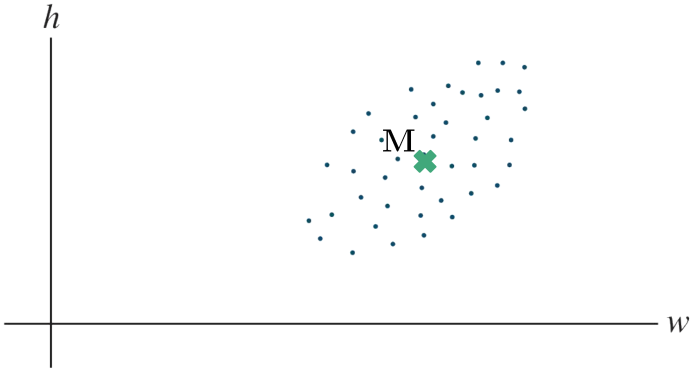


Diagrama de dispersión

Ejemplo PCA: media y covarianza

- Como preparación al análisis de componentes principales, definimos la **media muestral M** como:

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$$

$$B = \left[\, \hat{\mathbf{X}}_1 \;\; \hat{\mathbf{X}}_2 \;\; \cdots \;\; \hat{\mathbf{X}}_N \,\right]$$

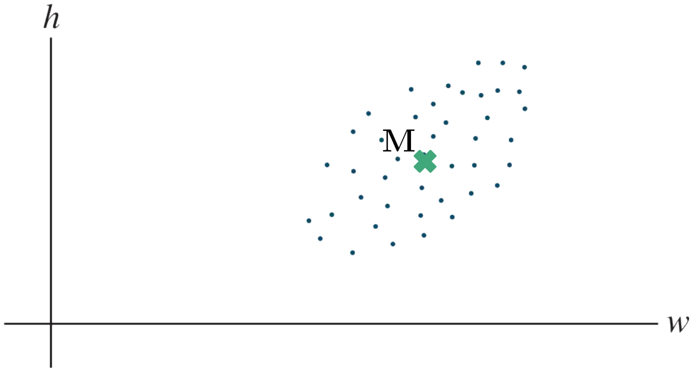


Diagrama de dispersión

Ejemplo PCA: media y covarianza

- Como preparación al análisis de componentes principales, definimos la **media muestral M** como:

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$$

$$B = \left[\, \hat{\mathbf{X}}_1 \;\; \hat{\mathbf{X}}_2 \;\; \cdots \;\; \hat{\mathbf{X}}_N \,\right]$$

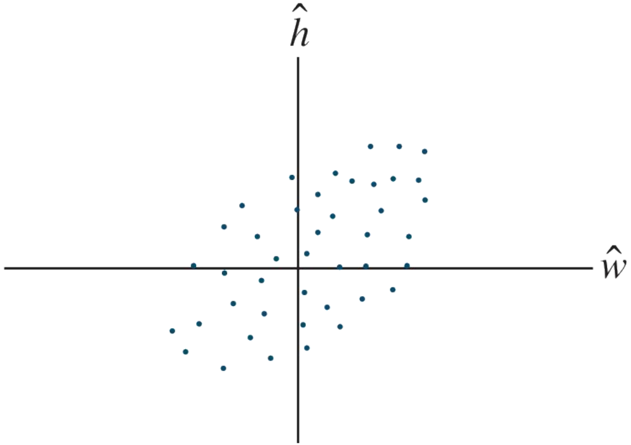


Diagrama de dispersión

Ejemplo PCA: media y covarianza

- Además, definimos la **matriz de covarianza** (muestral) como:

$$S = \frac{1}{N-1} B B^T$$

Notar que BB^T es semidefinida positiva, y por lo tanto, S también lo es.

Ejemplo PCA: media y covarianza

- Se realizan tres mediciones en cada uno de los cuatro individuos de una muestra aleatoria de una población. Calcular la media muestra y la matriz de covarianza para los siguientes vectores de observaciones:

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix}, \quad \mathbf{X}_4 = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix}$$

Ejemplo PCA: media y covarianza

- Se realizan tres mediciones en cada uno de los cuatro individuos de una muestra aleatoria de una población. Calcular la media muestra y la matriz de covarianza para los siguientes vectores de observaciones:

$$\mathbf{M} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix} \qquad S = \begin{bmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{bmatrix}$$

Ejemplo PCA: media y covarianza

- Se realizan tres mediciones en cada uno de los cuatro individuos de una muestra aleatoria de una población. Calcular la media muestra y la matriz de covarianza para los siguientes vectores de observaciones:

$$\mathbf{M} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix} \qquad S = \begin{bmatrix} \underline{10} & 6 & 0 \\ 6 & \underline{8} & -8 \\ 0 & -8 & \underline{32} \end{bmatrix}$$

Las entradas S_{jj} son la varianza de x_j

Ejemplo PCA: media y covarianza

- Se realizan tres mediciones en cada uno de los cuatro individuos de una muestra aleatoria de una población. Calcular la media muestra y la matriz de covarianza para los siguientes vectores de observaciones:

$$\mathbf{M} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix} \qquad S = \begin{bmatrix} \underline{10} & 6 & 0 \\ 6 & \underline{8} & -8 \\ 0 & -8 & \underline{32} \end{bmatrix}$$

$$\{ \text{ varianza total } \} = \operatorname{tr}(S)$$

Ejemplo PCA: media y covarianza

- Se realizan tres mediciones en cada uno de los cuatro individuos de una muestra aleatoria de una población. Calcular la media muestra y la matriz de covarianza para los siguientes vectores de observaciones:

$$\mathbf{M} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix} \qquad S = \begin{bmatrix} 10 & \underline{6} & \underline{0} \\ \underline{6} & 8 & \underline{-8} \\ \underline{0} & \underline{-8} & 32 \end{bmatrix}$$

Las entradas S_{ij} con $i \neq j$ son la **covarianza** de x_i y x_j

Análisis de componentes principales

- El análisis de datos multivariados en $\mathbf{X}_1, \ldots, \mathbf{X}_N$ se simplifica notablemente cuando la mayoría de las variables $x_1, \ldots, x_p$ (o todas) **no están correlacionadas**.
- Esto es lo mismo que decir que la matriz de covarianza de $\mathbf{X}_1, \ldots, \mathbf{X}_N$ es diagonal (o casi diagonal).

Análisis de componentes principales

- Suponiendo que la matriz $[\mathbf{X}_1 \ \ldots \ \mathbf{X}_N]$ ya tiene media 0, el objetivo del análisis de componentes principales es encontrar una matriz ortogonal $P = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_p]$ de $p \times p$, que determine un cambio de variable $\mathbf{X} = P\mathbf{Y}$:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

con la propiedad de que las nuevas variables $y_1, \ldots, y_p$ no estén correlacionadas.

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$S_{\mathbf{Y}} = \frac{1}{N-1}\left[\, \mathbf{Y}_1 \ \dots \ \mathbf{Y}_N \,\right]\left[\, \mathbf{Y}_1 \ \dots \ \mathbf{Y}_N \,\right]^T$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$S_{\mathbf{Y}} = \frac{1}{N-1}\underbrace{\left[\mathbf{Y}_1 \ \ldots \ \mathbf{Y}_N\right]}\underbrace{\left[\mathbf{Y}_1 \ \ldots \ \mathbf{Y}_N\right]}^T$$

$$\mathbf{X}_k = P\mathbf{Y}_k$$
$$P^{-1}\mathbf{X}_k = \mathbf{Y}_k$$
$$P^{T}\mathbf{X}_k = \mathbf{Y}_k$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$S_{\mathbf{Y}} = \frac{1}{N-1} \underbrace{\left[\, \mathbf{Y}_1 \, \ldots \, \mathbf{Y}_N \,\right]} \underbrace{\left[\, \mathbf{Y}_1 \, \ldots \, \mathbf{Y}_N \,\right]^T}$$

$$\mathbf{X}_k = P\mathbf{Y}_k$$

$$P^{-1}\mathbf{X}_k = \mathbf{Y}_k$$

$$P^T\mathbf{X}_k = \mathbf{Y}_k \longrightarrow P^T \left[\, \mathbf{X}_1 \, \ldots \, \mathbf{X}_N \,\right]$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$
\begin{aligned}
S_{\mathbf{Y}} &= \frac{1}{N-1} \underbrace{\left[\,\mathbf{Y}_1 \;\ldots\; \mathbf{Y}_N\,\right]}_{} \underbrace{\left[\,\mathbf{Y}_1 \;\ldots\; \mathbf{Y}_N\,\right]^T}_{} \\
&= \frac{1}{N-1} P^T \left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right] \left(P^T \left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right]\right)^T
\end{aligned}
$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$
\begin{aligned}
S_{\mathbf{Y}} &= \frac{1}{N-1}\left[\,\mathbf{Y}_1 \;\ldots\; \mathbf{Y}_N\,\right]\left[\,\mathbf{Y}_1 \;\ldots\; \mathbf{Y}_N\,\right]^T \\
&= \frac{1}{N-1}P^T\left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right]\left(P^T\left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right]\right)^T \\
&= P^T\left(\frac{1}{N-1}\left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right]\left[\,\mathbf{X}_1 \;\ldots\; \mathbf{X}_N\,\right]^T\right)P
\end{aligned}
$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal:

$$
\begin{aligned}
S_{\mathbf{Y}} &= \frac{1}{N-1}\left[\mathbf{Y}_1 \ \dots \ \mathbf{Y}_N\right]\left[\mathbf{Y}_1 \ \dots \ \mathbf{Y}_N\right]^T \\
&= \frac{1}{N-1}P^T\left[\mathbf{X}_1 \ \dots \ \mathbf{X}_N\right]\left(P^T\left[\mathbf{X}_1 \ \dots \ \mathbf{X}_N\right]\right)^T \\
&= P^T\left(\frac{1}{N-1}\left[\mathbf{X}_1 \ \dots \ \mathbf{X}_N\right]\left[\mathbf{X}_1 \ \dots \ \mathbf{X}_N\right]^T\right)P \\
&= P^T S P \longrightarrow
\end{aligned}
$$

$S_{\mathbf{Y}}$ en términos de S (covarianza de $\mathbf{X}$) y P

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal.
- Elegimos P, de tal manera que $P^T S P$ sea una matriz diagonal D:

$$S = PDP^T \longrightarrow \text{Diagonalización de } S$$

Análisis de componentes principales

- Queremos $\mathbf{X} = P\mathbf{Y}$ tal que $S_{\mathbf{Y}}$ es diagonal.
- Elegimos P, de tal manera que $P^T S P$ sea una matriz diagonal D:

$$S = PDP^T \longrightarrow \text{Diagonalización de } S$$

$$\underbrace{P^T S P = D}_{S_{\mathbf{Y}} \text{ como matriz diagonal}}$$

Análisis de componentes principales

- Finalmente:
  - P es la matriz que diagonaliza a S, y sus columnas son los vectores propios unitarios $\mathbf{u}_1, \ldots, \mathbf{u}_p$ (llamados **componentes principales**) de la matriz de covarianza S.
  - D es una matriz diagonal (matriz de covarianza $S_{\mathbf{Y}}$) con los valores propios de S, $\lambda_1, \ldots, \lambda_p$ sobre la diagonal, ordenados de tal manera que $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$.

Análisis de componentes principales: ejemplo

- A partir de la siguiente tabla lista los pesos y las estaturas:

	#1	#2	#3	#4	#5
Peso (kg)	54	57	57	61	61
Altura (cm)	155	152	163	173	172

- Encontrar la matriz de covarianza de los datos
- Hacer una análisis de componentes principales de los datos, para encontrar **un sólo índice de tamaño** que explique la mayor parte de la variación en los datos.

Análisis de componentes principales: ejemplo

- Computamos y restamos la media muestral a los datos:

$$\mathbf{M} = \begin{bmatrix} 58 \\ 163 \end{bmatrix} \qquad B = \begin{bmatrix} -4 & -1 & -1 & 3 & 3 \\ -8 & -11 & 0 & 10 & 9 \end{bmatrix}$$

Análisis de componentes principales: ejemplo

- Calculamos la matriz de covarianza de la muestra:

$$S = \frac{1}{5-1}\begin{bmatrix} -4 & -1 & -1 & 3 & 3 \\ -8 & -11 & 0 & 10 & 9 \end{bmatrix}\begin{bmatrix} -4 & -8 \\ -1 & -11 \\ -1 & 0 \\ 3 & 10 \\ 3 & 9 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 36 & 100 \\ 100 & 366 \end{bmatrix} = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix}$$

Análisis de componentes principales: ejemplo

- Calculamos autovalores y autovectores unitarios de S:

$$S = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix} \longrightarrow \begin{array}{l} \lambda_1 = 98.484 \\ \lambda_2 = 2.015 \end{array} \longrightarrow \mathbf{u}_1 = \begin{bmatrix} .269 \\ .963 \end{bmatrix}$$

Análisis de componentes principales: ejemplo

- Calculamos autovalores y autovectores unitarios de S:

$$S = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix} \longrightarrow \begin{array}{l} \lambda_1 = 98.484 \\ \lambda_2 = 2.015 \end{array} \longrightarrow \mathbf{u}_1 = \begin{bmatrix} .269 \\ .963 \end{bmatrix}$$

- El índice de tamaño lo podemos calcular usando la primera componente principal:

$$y_1 = .269\underset{\text{Peso}}{\hat{p}} + .963\underset{\text{Altura}}{\hat{a}}$$

Análisis de componentes principales: ejemplo

- ¿Qué porcentaje del total de la varianza explica este índice?
  - Varianza total:

$$tr(S) = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix} = 9.0 + 91.5 = 100.5$$

Análisis de componentes principales: ejemplo

- ¿Qué porcentaje del total de la varianza explica este índice?
  - Varianza total:

$$tr(S) = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix} = 9.0 + 91.5 = 100.5$$

  - Varianza del índice: $\lambda_1 = 98.484$

Análisis de componentes principales: ejemplo

- ¿Qué porcentaje del total de la varianza explica este índice?
  - Varianza total:

$$tr(S) = \begin{bmatrix} 9.0 & 25.0 \\ 25.0 & 91.5 \end{bmatrix} = 9.0 + 91.5 = 100.5$$

  - Varianza del índice: $\lambda_1 = 98.484$
  - Porcentaje del total: 97.9%

Otras aplicaciones

- Reducción de Dimensionalidad en Aprendizaje Automático
- Genómica y Bioinformática
- Finanzas
- Análisis de Datos de Encuestas
- Análisis de Imágenes