UC | Chile

Algebra Lineal Aplicada para Ciencia de Datos

Clase 6. Factorización QR

Introducción

Introducción y motivación

Factorización $A = QR$ es una de los conceptos mas importantes en Algebra Lineal. En esta clase responderemos primero la pregunta fundamental de como decidir si un conjunto de vectores en linealmente independiente. Luego avanzaremos a mostrar la factorización y sus aplicaciones para resolver sistemas lineales, proyecciones ortogonales, pseudoinversa, y mas adelante problemas de mínimos cuadrados.

Tema 1: Ortogonalización

Matrices ortogonales

Definición: Una colección de vectores $a_1, ..., a_k$ es ortogonal o mutuamente ortogonal si

$$a_i \cdot a_j = a_i^\top a_j = 0, \quad \text{para } 1 \leq i, j \leq k, \, i \neq j.$$

Si además tenemos que $\|a_i\|_2^2 = a_i^\top a_i = 1$, entonces los vectores se dicen ortonormales, es decir

$$a_i \cdot a_j = a_i^\top a_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Ejemplos: de vectores ortogonales

- Los vectores canónicos unitarios son ortonormales

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Los siguientes vectores son ortonormales

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Observación

- Si tenemos un conjunto de vectores **ortonormales** entonces tenemos que estos vectores son inmediatamente **linealmente independientes**.
  En efecto, si $a_1, ..., a_k$ son ortonormales y existen constantes $\beta_1, ..., \beta_k$ tales que

  $$\beta_1 a_1 + ... + \beta_k a_k = 0 \quad \Longrightarrow \quad 0 = a_i^\top(\beta_1 a_1 + ... + \beta_k a_k) = \beta_i \|a_i\|^2 \quad \Longrightarrow a_i = 0.$$

- Si $a_1, ..., a_k$ son vectores ortonormales y $x = \beta_1 a_1 + ... + \beta_k a_k$, entonces las constantes están dadas por $\beta_i = a_i^\top x$, $1 \leq i \leq k$.
- Un conjunto de $n-$ vectores ortonormales $a_1, ...a_n$ forman una base y se dicen base ortonormal.

Ejemplo base ortonormal

Sea el vector $x \in \mathbb{R}^3$, dado por $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, y la base ortonormal

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Entonces, describimos x en términos de la base ortonormal por

$$x = (x^\top a_1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (x^\top a_2) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (x^\top a_3) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Ejemplo base ortonormal

Sea el vector $x \in \mathbb{R}^3$, dado por $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, y la base ortonormal

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Entonces, describimos x en términos de la base ortonormal por

$$x = (-3) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (\frac{3}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (-\frac{1}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Pregunta

¿Como podemos saber si una lista de vectores son linealmente independentes?

Proceso de ortogonalización de Gram-Schmidt

Algorithm 1: Gram - Schmidt (clásico)

Data: Vectores de $\mathbb{R}^n$, $a_1, ..., a_k$.

Result: Vectores ortonormales de $\mathbb{R}^n$, $q_1, ..., q_\ell$, $\ell \leq k$.

```
for i = 1 : k do
    q̃_i = a_i − (q_1^⊤ a_i) q_1 − ... (q_{i−1}^⊤ a_i) q_{i−1};   // Ortogonalización
    if q̃_i = 0 then
        break ;                                                  // dependencia lineal
    else
        q_i = q̃_i / ‖q̃_i‖_2 ;                                    // normalización
```

for $i = 1 : k$ **do**

- $\tilde{q}_i = a_i - (q_1^\top a_i)q_1 - \ldots (q_{i-1}^\top a_i)q_{i-1}$; `// Ortogonalización`
- **if** $\tilde{q}_i = 0$ **then**
  - break ; `// dependencia lineal`
- **else**
  - $q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$; `// normalización`

Ejemplo de ortogonalización

Considere los vectores $a_1 = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$

El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \begin{bmatrix} 0.2425 \\ 0.9701 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 0.9701 \\ -0.2425 \end{bmatrix}.$$

Ilustración gráfica

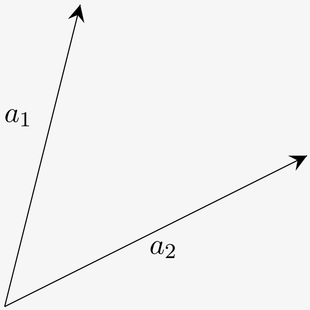


Ilustración gráfica

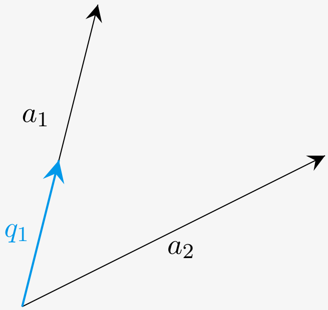


Ilustración gráfica

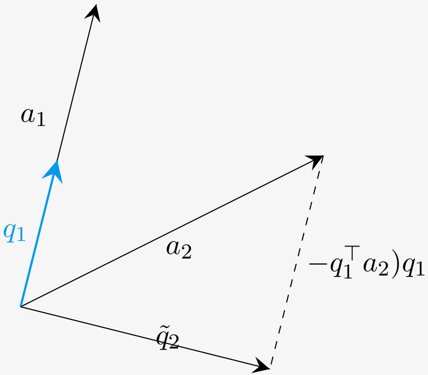
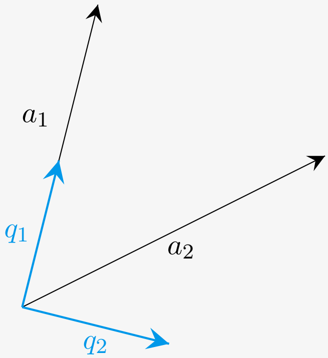


Ilustración gráfica



Ejemplo de ortogonalización

Considere los vectores $$a_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}.$$

El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad q_2 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \quad q_3 = \sqrt{14} \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Observaciones

1. Si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt se completa entonces los vectores son linealmente independientes.
2. Si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt no se completa, o termina prematuramente en la iteración j, entonces el vector a_j es una combinación lineal de los vectores $q_1, ..., q_{j-1}$ (o también de $a_1, ..., a_{j-1}$).
3. **Algoritmo de Gram-Schmidt Modificado**. En implementaciones una version modificada del algoritmo, pero equivalente matemáticamente, se considera debido a que posee propiedades de errores de redondeo superiores (**ver Tutorial 5**).

Tema 2: Factorización QR

Definición de matriz ortogonal

Una matrix $A \in \mathbb{R}^{m \times n}$ se dice **ortogonal** si satisface que

$$A^\top A = I \in \mathbb{R}^{m \times m}.$$

Observe que las columnas de matrices ortogonales forman una base ortonormal.

Factorización QR

Dado un conjunto de vectores linealmente independientes $a_1, ..., a_n$, el proceso de Gram-Schmidt nos entrega los vectores ortonormales $q_1, ..., q_n$ y podemos escribir

$$\begin{aligned} a_i &= (q_1^\top a_i)q_1 + ... + (q_{i-1}^\top a_i)_{i-1} + +(q_i^\top a_i)q_i \\ &= R_{1i}q_1 + ... + R_{i-1i}q_{i-1} + R_{ii}q_i \end{aligned}$$

Entonces, si $A \in \mathbb{R}^{m\times n}$ con columnas linealmente independientes, usamos el proceso de ortogonalización de Gram-Schmidt para las columnas de A

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} R = QR$$

donde $R \in \mathbb{R}^{n\times n}$ es triangular superior con elementos en la diagonal no cero y $Q \in \mathbb{R}^{m\times n}$ es ortogonal.

Ejemplo de factorización QR.

Se la matriz

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -2 \\ -1 & 2 & 3 \end{bmatrix}$$

Entonces, la factorización QR de A es

$$A = \underbrace{\begin{bmatrix} 1/\sqrt{3} & 4/\sqrt{42} & 2/\sqrt{14} \\ 1/\sqrt{3} & 1/\sqrt{42} & -3/\sqrt{14} \\ -1/\sqrt{3} & 5/\sqrt{42} & -1/\sqrt{14} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{3} & -1/\sqrt{3} & -\sqrt{3} \\ 0 & \sqrt{14}/\sqrt{3} & \sqrt{21}/\sqrt{2} \\ 0 & 0 & \sqrt{7}/\sqrt{2} \end{bmatrix}}_{R}$$

Resolver ecuaciones lineales con QR

Sea $A \in \mathbb{R}^{n\times n}$ una matriz no singular y sea $n \in \mathbb{R}^n$. Entonces

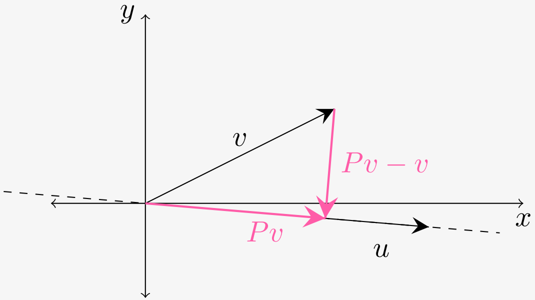
$$Ax = b \quad \Longleftrightarrow \quad (QR)x = b \quad \Longleftrightarrow \quad Rx = Q^\top b.$$

Algorithm 2: Solver de ecuaciones lineales con factorización QR

Data: $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$
Result: $x \in \mathbb{R}^n$.
Calcular factorización $A = QR$;
Calcular $Q^\top b$;
Resolver sistema triangular $Rx = Q^\top b$;

Matriz de proyección

Definición: Una proyecccion es una matriz cuadrada P que satisface $P^2 = P$. Además, decimos que una proyección es ortogonal si $P^\top = P$



Proyección ortogonal

Observación: Si Q es una matriz ortogonal alta, entonces $P = QQ^\top$ satisface que:

$$P^2 = (QQ^\top)(QQ^\top) = QQ^\top = P$$
$$P^\top = (QQ^\top)^\top = QQ^\top = P$$

Tema 3: Pseudoinversa

Matriz de Gram

Definición. Sea $A \in \mathbb{R}^{m\times n}$. se define la matriz de Gramm asociada como la matriz cuadrada $A^\top A$.
Observación. Note que A tiene columnas linealmente independientes si y sólo sí su matriz de Gram es invertible. En efecto,

$$
\begin{aligned}
(A^\top A)x = 0 \quad &\longrightarrow \quad 0 = x^\top (A^\top A)x = x^\top A^\top Ax = \|Ax\|^2 \\
&\longrightarrow \quad Ax = 0 \\
&\longrightarrow \quad x = 0
\end{aligned}
$$

Para el recíproco observe que si existe $x \neq 0$ tal que $Ax = 0$, entonces $A^\top Ax = 0$, lo que implica que la matriz de Gram no es invertible.

Definición de Pseudoinversa

Sea $A \in \mathbb{R}^{m \times n}$.

- Si A tiene columnas linealmente independientes entonces la matriz $(A^\top A)^{-1} A^\top$ es una inversa por la izquierda de A (matrices altas o cuadradas.) La peudoinversa se define entonces

$$A^\dagger = (A^\top A)^{-1} A^\top$$

- Si A tiene filas linealmente independientes entonces la matriz $A^\top (A A^\top)^{-1}$ es una inversa por la derecha de A (matrices anchas o cuadradas.) La peudoinversa se define entonces

$$A^\dagger = A^\top (A A^\top)^{-1}$$

Pseudoinvera por factorización QR

Suponga que $A \in \mathbb{R}^{m \times n}$, y sus columnas son linealmente independientes, y sea su factorización $A = QR$. Entonces:

$$A^\top A = (QR)^\top (QR) = R^\top R$$

y así su pseudoinversa queda:

$$A^\dagger = (A^\top A)^{-1} A = (R^\top R)^{-1} (QR)^\top = R^{-1} Q^\top$$

Resolviendo sistemas sobredeterminados

Considere el sistema sobre determinado

$$Ax = b \quad \iff \quad \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

La factorización QR de A es

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -0.5883 & 0.4576 \\ 0.7845 & 0.5230 \\ 0.1961 & -0.7191 \end{bmatrix} \begin{bmatrix} 5.0990 & 7.2563 \\ 0 & 0.5883 \end{bmatrix}$$

La solución usando la pseudoinversa es (no necesariamente solución del sistema lineal)

$$x = A^{\dagger}b = R^{-1}Q^{\top} = \begin{bmatrix} -1.2222 & -1.1111 & 1.7778 \\ 0.7778 & 0.8889 & -1.2222 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

UC | Chile