

Clase 33. Inversas izquierda, derecha y Pseudoinversa de una Matriz.

Curso ‘Linear Algebra’ del MIT.

Resumen

Si bien la inversa ideal de una matriz solo podemos encontrarla en una cuadrada, igual es posible encontrar otras matrices (rectangulares o cuadradas singulares) que tienen elementos de estas últimas. Se conocen como **inversas generalizadas** y en esta clase veremos dos tipos de ellas: las **inversas unilaterales** y la **pseudoinversa**.

1. Inversas unilaterales.

Supongamos que A es una matriz de $n \times m$, con $n \neq m$. En principio no es invertible puesto que no es cuadrada. No obstante, dependiendo de su **rango** (*rank*) podemos encontrar la inversa de un lado conocidas como **Inversas unilaterales**.

Existen dos inversas unilaterales de una matriz: La inversa izquierda y la inversa derecha. Reciben estos nombres porque con ellas es posible obtener la matriz I multiplicando a A solo por **un lado**. Es decir, se cumple una parte de $A^{-1}A = AA^{-1} = I$.

1.1 Inversa izquierda de una matriz.

La inversa izquierda C de una matriz $A \in \mathbb{R}^{n \times m}$, con $n \neq m$, es aquella donde:

$$C \cdot A = I, \text{ pero } A \cdot C \neq I$$

Para encontrar a C usamos a la matriz $A^T A$. En la Clase 30 (págs 12-13) demostramos que

$$\text{rango}(A) = \text{rango}(A^T A)$$

Si el $\text{rango}(A) = m$, entonces

$$\text{rango}(A^T A) = m$$

Es decir, $A^T A$ es de **rango completo** cuando A es de **rango columna completo**. Esto implica que:

$$(A^T A)^{-1}(A^T A) = I$$

La matriz C se obtiene ordenando el lado izquierdo de la ecuación anterior como

$$\left[(A^T A)^{-1} A^T\right] A = I$$

donde

$$C = (A^T A)^{-1} A^T$$

es la **matriz inversa izquierda** de A si y solo si el $\text{rango}(A) = m$.

Ahora, si bien $AC \neq I$, reemplacemos a C en el producto matricial.

$$AC = A \left[(A^T A)^{-1} A^T\right] = P$$

Como vemos, AC es la **matriz de proyección ortogonal** sobre el $C(A)$.¹

La matriz de proyección ortogonal es una transformación lineal que entrega las **mejores aproximaciones** a los vectores de entrada. Aquello pasaría a ser una igualdad si $P = I$, pero no lo es porque C no es la inversa derecha de A .

En ese sentido, C en AC opera como una “**inversa bilateral aproximada**” o *pseudoinversa* de A . Más adelante profundizaremos en este concepto.

1.2 Inversa derecha de una matriz.

La otra inversa unilateral de una matriz $A \in \mathbb{R}^{n \times m}$ se conoce como **inversa derecha** y es cualquier matriz $D \in \mathbb{R}^{m \times n}$ donde se cumple que

$$AD = I, \text{ pero } DA \neq I$$

Para conocer a D calculamos la matriz $AA^T \in \mathbb{R}^{n \times n}$. Usando el teorema de nulidad podemos demostrar que

$$\text{rango}(A^T) = \text{rango}(AA^T)$$

¹Más detalles en los apuntes de las clases 15 y 16.

Como el $\text{rango}(A^T) = \text{rango}(A)$, entonces

$$\text{rango}(A) = \text{rango}(AA^T)$$

Esta igualdad es importante porque para buscar a la matriz D necesitamos que A sea de **rango fila completo**. Es decir, que

$$\text{rango}(A) = n$$

Lo anterior implica que AA^T es de **rango completo**, ya que

$$\text{rango}(AA^T) = n$$

En consecuencia, AA^T tiene una **inversa**, cumpliéndose la igualdad

$$(AA^T)(AA^T)^{-1} = I$$

Al reordenar los paréntesis de las matrices del lado izquierdo obtenemos que

$$A \left[A^T (AA^T)^{-1} \right] = I \iff \text{rango}(A) = n$$

donde

$$D = A^T (AA^T)^{-1}$$

es la **inversa derecha** de A .

Por otra parte, veamos que $DA \neq I$ es la **matriz de proyección ortogonal** sobre el $C(A^T)$.

$$DA = A^T (AA^T)^{-1} A = P$$

En ese sentido, D en DA también opera como una “inversa aproximada” o **pseudoinversa** de A . Profundicemos en la siguiente sección sobre ella.

2. Pseudoinversa de una matriz.

En la sección anterior vimos que cuando multiplicamos a las inversas izquierda y derecha de A , C y D , en sentido contrario obtenemos aquellas que permiten proyectar ortogonalmente

sobre los subespacios $C(A)$ y $C(A^T)$.

$$AC = P \quad DA = P$$

En ambos casos, señalamos que C y D operan como matrices **pseudoinversas** de A .

Las pseudoinversas, o mejor conocidas como **Inversas Generalizadas de Moore-Penrose**, son transformaciones lineales similares a las inversas que se caracterizan por ser aplicables a matrices rectangulares o cuadradas singulares. Suelen denotarse como A^\dagger o A^+ .

En ese sentido, si el $\text{rango}(A) = m$, entonces su pseudoinversa es:

$$A^+ = (A^T A)^{-1} A^T = C$$

Mientras que si el $\text{rango}(A) = n$, entonces:

$$A^+ = A^T (A A^T)^{-1} = D$$

donde, a su vez, C y D son las inversas izquierda y derecha de A , respectivamente.

Si la matriz A **no es rango columna ni fila completo**, la manera más usada para calcular su pseudoinversa es **descomponiéndola en valores singulares**. Su fórmula proviene de aplicarla a su inversa.

Consideremos la SVD de una matriz $A \in \mathbb{R}^{n \times m}$.

$$A = U \Sigma V^T$$

Luego, tomemos la inversa de las matrices de ambos lados de la igualdad.

$$A^{-1} = (U \Sigma V^T)^{-1} = (V^T)^{-1} \Sigma^{-1} U^{-1}$$

Como U y V son matrices ortogonales, entonces $U^{-1} = U^T$ y $V^{-1} = V^T$. Por lo tanto,

$$A^{-1} = V \Sigma^{-1} U^T$$

Al reemplazar a las matrices inversas por las pseudoinversas, obtenemos a A^+ .

$$A^+ = V \Sigma^+ U^T$$

donde

$$\Sigma^+ = (\Sigma^{-1})^T$$

Debemos tener en cuenta que si A es singular, en estricto rigor no existe Σ^{-1} . La idea de la igualdad de arriba es **usar las operaciones para buscar a Σ^{-1}** . Cuando la transponemos y calculamos los productos con V y U^T , obtenemos a la pseudoinversa de A .

Por otra parte, veamos que si A es de $n \times m$ con $n = m$, entonces

$$\Sigma^+ = (\Sigma^{-1})^T = \Sigma^{-1},$$

ya que la transpuesta de una matriz **cuadrada** diagonal siempre resulta en ella misma.

Las primeras r entradas de la diagonal de Σ^+ siempre serán iguales a $1/\sigma_i$ (el resto serán ceros).

Recordemos que $A = U\Sigma V^T$ podemos expresarla como

$$A\vec{v}_i = \sigma_i\vec{u}_i$$

Es decir, en la SVD estamos aplicando la transformación lineal $\vec{v}_i \mapsto A\vec{v}_i = \sigma_i\vec{u}_i$, donde el dominio es el $C(A^T)$ y el codominio el $C(A)$. Si queremos intentar revertir este proceso, podemos multiplicar a la izquierda de ambos lados por A^+ .

$$A^+A\vec{v}_i = \sigma_i A^+\vec{u}_i$$

Si $A^+ = A^{-1}$, entonces obtenemos devuelta al vector \vec{v}_i .

$$\vec{v}_i = \sigma_i A^{-1}\vec{u}_i$$

En cambio, si $A^+ \neq A^{-1}$, obtenemos un vector que es la **proyección ortogonal** de \vec{v}_i sobre el $C(A^T)$, que denotaremos como \vec{x}_i^+ .

$$A^+A\vec{v}_i = \sigma_i A^+\vec{u}_i = \vec{x}_i^+ \neq \vec{v}_i$$

Donde A^+A es la **matriz de proyección** sobre el $C(A^T)$.

De modo similar, si $A^+ \neq A^{-1}$, AA^+ es la **matriz de proyección** sobre el $C(A)$ que entrega una aproximación de \vec{u}_i transformado previamente por A^+ en $\vec{u} \mapsto A^+\vec{u} = (1/\sigma_i)\vec{v}_i$. Esta proviene de $A^+ = V\Sigma^+U^T$.

2.1 Aplicación en el método de mínimos cuadrados ordinarios.

En estadística, es habitual evaluar la relación lineal de una variable aleatoria con respecto a otra o a más de una mediante un modelo del tipo

$$x_2(x_1) = \beta_0 + x_1\beta_1$$

Este modelo se aplica a cada observación y se construye un sistema de ecuaciones lineales para obtener los valores de β_0 y β_1 .

El problema de este proceso, es que es común terminar con un **sistema sobredeterminado** porque siempre opta por tener más observaciones para que la estimación de la relación sea lo más precisa posible. Como sabemos, esto nos lleva a dos escenarios:

1. Que el sistema tenga una solución.
2. Que el sistema no tenga solución (i.e, que sea **inconsistente**).

La mayoría de las veces se termina con un sistema sobredeterminado inconsistente.

Representemos este sistema como

$$A\vec{x} = \vec{b}$$

donde \vec{b} contiene los valores de x_2 en función de x_1 , \vec{x} a β_0 y β_1 , mientras que en A están los coeficientes del sistema.

Un sistema sobredeterminado es inconsistente cuando \vec{b} no es una combinación lineal de las columnas de A dada por $A\vec{x}$. Es decir, cuando \vec{b} no es miembro del $C(A)$. En dicho caso, en estadística se opta usar el método de **Mínimos Cuadrados Ordinarios** (OLS en inglés) para encontrar un modelo que se aproxime (o ajuste) mejor a los datos.

El método OLS consiste en buscar **estimaciones** de β_0 y β_1 que **minimicen** la suma de los **errores cuadráticos** del modelo lineal.

En la clase 16 estudiamos una manera de resolver el método OLS. Buscamos un vector \hat{x} que contiene a las estimaciones de β_0 y β_1 **proyectando ortogonalmente** a \vec{b} sobre el $C(A)$ mediante el sistema

$$A\hat{x} = \vec{b}$$

El objetivo fue buscar la **matriz de proyección ortogonal** P que realice la transformación $\vec{b} \mapsto P\vec{b} = \hat{x}$. Para ello, multiplicamos la igualdad de arriba por A^T por la izquierda.

$$A^T A\hat{x} = A^T \vec{b}$$

Si A es de **rango columna completo**, entonces $A^T A$ es de **rango completo**, implicando que $(A^T A)^{-1}$ existe. Al multiplicarla en la ecuación, obtenemos la matriz de proyección que nos permite encontrar a \hat{x} .

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}; \text{ donde } (A^T A)^{-1} A^T = P$$

Otra manera muy común de encontrar a la matriz de proyección ortogonal P , es usando a la **pseudoinversa** de A , A^+ , derivada de la **versión reducida** de la **SVD**. Volvamos al sistema

$$A\hat{x} = \vec{b}$$

En este caso, la idea es multiplicar a la izquierda de ambos lados esta ecuación por A^+ .

$$A^+ A \hat{x} = A^+ \vec{b}$$

Luego, usamos las igualdades $A = U\Sigma V^T$ y $A^+ = V\Sigma^+ U^T$ y las reemplazamos en la ecuación.

$$(V\Sigma^+ U^T)(U\Sigma V^T)\hat{x} = (V\Sigma^+ U^T)\vec{b}$$

Como $U^T U = I$, entonces:

$$V\Sigma^+ \Sigma V^T \hat{x} = (V\Sigma^+ U^T)\vec{b}$$

Debido a que estamos usando la versión reducida de la SVD, si A es de **rango columna completo**, entonces $\Sigma^+ \Sigma = I$ ya que $\Sigma \in \mathbb{R}^{r \times r}$. Esto conlleva a que el lado izquierdo de la ecuación sea $VV^T \hat{x}$. Como $VV^T = I$, en consecuencia:

$$\hat{x} = (V\Sigma^+ U^T)\vec{b} = A^+ \vec{b}$$

donde A^+ es la **matriz de proyección ortogonal**.

De este proceso podemos sacar una buena conclusión: Cuando A es de rango columna completo y A^+ proviene de la versión reducida de la SVD, entonces:

$$A^+ A = I$$