Ejemplos de Normas Matriciales

Alan Reyes-Figueroa

Métodos Numéricos II (Aula 02) 06.julio.2023

Ejemplo

Ejemplo: Calcular la norma 1 inducida de $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & 1 \end{pmatrix}$.

En este caso, tenemos tres vectores columna

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Las normas 1 de estos vectores columna son, respectivamente,

$$||\mathbf{a}_1||_1 = |1| + |0| = 1, \quad ||\mathbf{a}_2||_1 = |2| + |-3| = 5, \quad ||\mathbf{a}_3||_1 = |3| + |1| = 4.$$

Finalmente, el máximo de estas normas 1 vectoriales es 6, y tenemos

$$||A||_1 = \max_{1 \leq j \leq 3} ||\mathbf{a}_j||_1 = \max\{1, 5, 4\} = 5.$$

Ejemplo

Ejemplo: Calcular la norma ∞ inducida de $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & 1 \end{pmatrix}$.

En este caso, tenemos dos vectores fila

$$\mathbf{a}_1^T = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \quad \mathbf{a}_2^T = \begin{pmatrix} 0 & -3 & 1 \end{pmatrix}.$$

Las normas 1 de estos vectores fila son, respectivamente,

$$||\mathbf{a}_1^T||_1 = |1| + |2| + |3| = 6, \quad ||\mathbf{a}_2^T||_1 = |0| + |-3| + |1| = 4.$$

Finalmente, el máximo de estas normas 1 vectoriales es 6, y tenemos

$$||A||_\infty = \max_{1 \leq i \leq 2} ||\mathbf{a}_i^T||_1 = \max\{6, 4\} = 6.$$

Norma 2 Inducida

Teorema

Sea $A \in \mathbb{R}^{m\times n}$ una matriz, y sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ los autovalores de $A^TA \in \mathbb{R}^{m\times m}$. Entonces

$$||A||_2 = \Big(\max_{1\leq i\leq m} |\lambda_i| \Big)^{1/2} = \sqrt{\max_{1\leq i\leq m} \lambda_i}.$$

Prueba: La prueba se dará más adelante.

Observe que:

- A^TA es una matriz cuadrada simétrica, portanto sus autovalores son siempre números reales.
- Casi siempre ocurre que A^TA es semi-definida positiva, de modo que sus autovalores son ≥ 0.
- $\max_{1\leq i\leq m} \lambda_i$ es el mayor autovalor de A^TA. Este se llama el **radio espectral** de A^TA.

Ejemplo: Calcular la norma 2 inducida para $A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$.

En este caso, comenzamos calculando la matriz A^TA

$$A^TA = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}.$$

Luego, el polinomio característico de A^TA es

$$\det(\lambda I - A^TA) = \lambda^2 - \mathrm{tr}(A^TA)\,\lambda + \det(A^TA) = \lambda^2 - 9\lambda + 4 = 0,$$

cuyas raíces son $\lambda_{\max} = \dfrac{9+\sqrt{65}}{2}$ y $\lambda_{\min} = \dfrac{9-\sqrt{65}}{2}$.

Luego, $||A||_2 = \sqrt{\lambda_{\max}} = \sqrt{\dfrac{9+\sqrt{65}}{2}} \approx 2.92080962648$.

Ejemplo

Ejemplo: Calcular la norma 2 inducida para $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & 1 \end{pmatrix}$.

Respuesta: $||A||_2 \approx 3.911$ (Ejercicio!)

Normas Matriciales

En general, no todas las normas matriciales son inducidas por una norma vectorial. Damos la siguiente definición general.

Definición

*Una **norma matricial** es una función* $||\cdot|| : \mathbb{R}^{m\times n} \to \mathbb{R}$ *que satisface:*

1. $||A|| \geq 0, \;\; \forall A \in \mathbb{R}^{m\times n}$.
2. $||A+B|| \leq ||A|| + ||B||, \forall A, B \in \mathbb{R}^{m\times n}$.
3. $||\alpha A|| = |\alpha| \; ||A||, \forall \alpha \in \mathbb{R}, \forall A \in \mathbb{R}^{m\times n}$.

La norma matricial más importante que no proviene de una norma vectorial es la **norma de Frobenius** o **norma de Hilbert-Schmidt**

$$||A||_F = \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \Big)^{1/2}. \tag{1}$$

Normas Matriciales

Si $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ denotan las columnas de A, entonces

$$||A||_F = \Big(\sum_{j=1}^{n} ||\mathbf{a}_j||_2^2 \Big)^{1/2}. \tag{2}$$

Proposición

Para $A \in \mathbb{R}^{m\times n}$, vale $||A||_F = \sqrt{\mathrm{tr}(A^TA)} = \sqrt{\mathrm{tr}(AA^T)}$.

<u>Prueba</u>: Si $A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix}$ son las columnas de A, $\mathbf{a}_j \in \mathbb{R}^m$, entonces

$$A^TA = \begin{pmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{pmatrix} = \left(\mathbf{a}_i^T\mathbf{a}_j\right).$$

Luego, $\mathrm{tr}(A^TA) = \sum_{j=1}^{n} \mathbf{a}_j^T\mathbf{a}_j = \sum_{j=1}^{n} ||\mathbf{a}_i||_2^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2 = ||A||_F^2$. $\square$

Ejemplo

Ejemplo: Calcular la norma 2 inducida para $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & 1 \end{pmatrix}$.

Tenemos que

$$\sum_{i=1}^{2}\sum_{j=1}^{3} a_{ij}^2 = 1^2 + 2^2 + 3^2 + 0^2 + (-3)^2 + 1^2 = 24.$$

Luego, $||A||_F = \sqrt{24} = 4.89897948$.

Otra manera de calcular la norma de Frobenius sería usando $||A||_F = \sqrt{\mathrm{tr}(A^TA)}$. Observe que

$$A^TA = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 13 & 3 \\ 3 & 3 & 10 \end{pmatrix} \qquad \Longrightarrow \qquad \mathrm{tr}(A^TA) = 1 + 13 + 10 = 24.$$

Portanto, $||A||_F = \sqrt{24} = 4.89897948$.

¿Para qué sirven las normas matriciales?

Tiene sentido entonces definir normas matriciales, para medir el "tamaño" de estas matrices.

Recordemos que en el contexto de métodos numéricos, las normas vectoriales (y matriciales nos sirven para calcular distancias entre vectores (o matrices).

$$\text{dist}(\mathbf{x}, \mathbf{y}) = ||\mathbf{y} - \mathbf{x}||_p.$$

y las distancias nos dan un mecanismo para medir el error de aproximación. Por ejemplo, si $\mathbf{x} \in \mathbb{R}^n$ es la solución de un problema, y tenemos un algoritmo numérico que nos devuelve una solución aproximada $\widehat{\mathbf{x}} \in \mathbb{R}^n$, entonces el error de aproximación es

$$\text{error de aproximación} = ||\mathbf{x} - \widehat{\mathbf{x}}||_p.$$

Lo mismo ocurre para un algoritmo que calcular matrices:

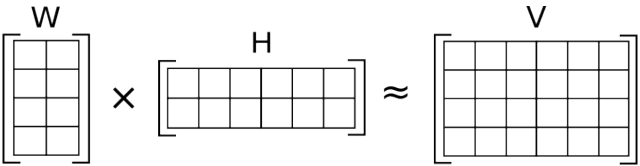
$$\text{error de aproximación} = ||A - \widehat{A}||_p.$$

Aplicación

La factoración de matrices es una estrategia muy usada para obtener variables lantentes, útiles para revelar información no observable dentro de una matriz de datos $\mathbb{X}$ (e.g. un sistema de recomendación).

Sea $\mathbb{X} \in \mathbb{R}^{n \times d}$ una matriz de datos. El objetivo es descomponer $\mathbb{X}$ como el producto de dos matrices con entradas no-negativas $W \in \mathbb{R}^{n \times r}$ y $H \in \mathbb{R}^{r \times d}$

$$\mathbb{X} = WH,$$



La solución de este algoritmo consiste en encontrar la combinación de matrices $\widehat{W}$ y $\widehat{H}$ que minimiza el error de aproximación $||\mathbb{X} - \widehat{W}\widehat{H}||_F$.