Descomposición SVD

Alan Reyes-Figueroa

Métodos Numéricos II

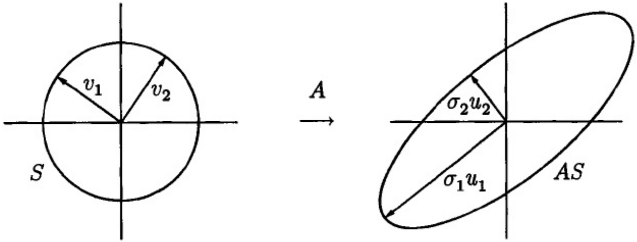
(Aula 04) 13.julio.2023

Descomposición SVD

La descomposición en valores singulares, SVD, está motivada por el siguiente hecho geométrico:

La imagen de la esfera unitaria bajo cualquier matriz en $\mathbb{R}^{m\times n}$ es una hiperelipse en $\mathbb{R}^m$, obtenida al "estirar" la esfera unitaria en $\mathbb{R}^m$ por algunos factores $\sigma_1, \ldots, \sigma_m$, (posiblemente cero) en algunas direcciones ortogonales $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^m$.

Por conveniencia, consideramos que los $\mathbf{u}_i$ son vectores unitarios, es decir, $||\mathbf{u}_i||_2 = 1$. Los vectores $\{\sigma_i \mathbf{u}_i\}$ son los semi-ejes principales de la hiperelipse, con longitudes $\sigma_1, \ldots, \sigma_m$. Si A tiene rango r, exactamente r de estas longitudes resultarán ser distintas de cero, y en particular, si $m > n$, como máximo n de ellas serán distinto de cero.



SVD de una matriz en $\mathbb{R}^{2\times 2}$.

Sea $S = S^{n-1}$ la esfera unitaria en $\mathbb{R}^n$, y tome cualquier matriz $A \in^{m\times n}$ con $m > n$. Por simplicidad, supongamos que A tiene rango completo n. La imagen AS es una hiperelipse en $\mathbb{R}^m$.

Definimos algunas propiedades de A en términos de la forma de AS.

Descomposición SVD

Primero, definimos los n **valores singulares** de A. Estos son las longitudes de los n semi-ejes principales de AS, escritos $\sigma_l \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$.

Los n **vectores singulares izquierdos** de A. Estos son los vectores unitarios $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n n\}$ orientados en las direcciones de los semi-ejes principales de AS, numerados para corresponder con los valores singulares. Así, $\sigma_i \mathbf{u}_i$ es el i-ésimo mayor semi-eje principal de AS.

Finalmente, los n **vectores singulares derechos** de A son los vectores unitarios $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset S$ que son las preimágenes de los semi-ejes principales de AS, numerados de modo que $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$, $i = 1, 2, \ldots, n$.

De lo anterior tenemos una relación entre los σ_i, los $\mathbf{u}_i$ y los $\mathbf{v}_i$:

Descomposición SVD

$$A\mathbf{v}_j = \sigma_j \mathbf{u}_j, \ \forall\, 1 \leq j \leq n. \tag{1}$$

Esta colección de ecuaciones vectoriales se puede expresar como una ecuación matricial, como

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} \\ v_1 \,\Big|\, v_2 \,\Big|\, \cdots \,\Big|\, v_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ u_1 \,\Big|\, u_2 \,\Big|\, \cdots \,\Big|\, u_n \\ \\ \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

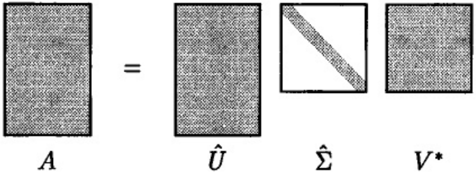
o de forma más compacta como $AV = \widehat{U}\widehat{\Sigma}$, con $\widehat{U} \in \mathbb{R}^{m\times n}$, $V, \widehat{\Sigma} \in \mathbb{R}^{n\times n}$. $V \in O(n)$ es ortogonal, mientras que $\widehat{U}$ posee columnas ortogonormales.

Como V es ortonormal, podemos escribir $A = \widehat{U}\widehat{\Sigma}V^T$.

Descomposición SVD

Definición

La factoración $A = \widehat{U}\widehat{\Sigma}V^T$ se llama la ***descomposición en valores singulares reducida*** *de A.*



Descomposición SVD reducida, (para $m \geq n$).

Introducimos una segunda factoriación SVD. La idea es la siguiente. Las columnas de $\widehat{U}$ son n vectores ortonormales en el espacio $\mathbb{R}^m$. A menos que $m = n$, no forman una base de $\mathbb{R}^m$, ni $\widehat{U}$ es una matriz ortogonal.

Descomposición SVD

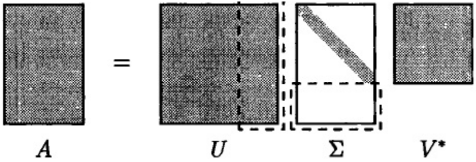
Adjuntamos $m - n$ columnas ortonormales adicionales, de modo que $\widehat{U}$ se extiende a una matriz ortogonal U. Si $\widehat{U}$ se reemplaza por U en (1), $\widehat{\Sigma}$ también cambia. Para que el producto permanezca inalterado, las últimas $m - n$ columnas de U deben multiplicarse por cero. En consecuencia, sea Σ la matriz $m \times n$ que consta de $\widehat{\Sigma}$ en el bloque $n \times n$ superior junto con $m - n$ filas de ceros a continuación.

Ahora tenemos una nueva factoración de la forma $AV = U\Sigma$ ó $A = U\Sigma V^T$, donde $\widehat{U} \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^n$ y $\Sigma \in \mathbb{R}^{m\times n}$. Además, $V \in O(n)$ y $U \in O(m)$ son matrices ortogonales.

Descomposición SVD

Definición

*La factoración $A = U\Sigma V^T$ se llama la **descomposición en valores singulares completa** de A.*



Descomposición SVD completa, (para $m \geq n$).

Habiendo descrito la SVD completa, ahora podemos descartar el supuesto inicial que A tiene rango completo. Si A es deficiente en rango, i.e. $\text{rank}(A) = r < \min\{m, n\}$, la factoración SVD completa aún es válida.

Descomposición SVD

Todo lo que cambia es que ahora no n sino solo r de los vectores singulares izquierdos de A están determinados por la geometría de la hiperelipse. Al construir la matriz ortogonal U, añadimos $m - r$ en lugar de $m - n$ columnas ortonormales adicionales. La matriz V también necesitará $n - r$ columnas ortonormales adicionales para extender las r columnas determinadas por la geometría.

La matriz Σ ahora tendrá r entradas diagonales positivas, con las $n - r$ restantes igual a cero. De la misma manera, la SVD reducida también tiene sentido para matrices A de rango deficiente. Se puede tomar $\widehat{U}$ como $m \times n$, con $\widehat{\Sigma}$ de dimensiones $n \times n$ con algunos ceros en la diagonal, o comprimir aún más la representación de modo que $\widehat{U}$ sea $m \times r$ y $\widehat{\Sigma}$ sea $r \times r$ y estrictamente positiva en la diagonal.

Teorema (Existencia y unicidad SVD)

Toda matriz $A \in \mathbb{C}^{m\times n}$ posee una descomposición en valores singulares $A = U^*$. Además, los valores singulares $\{\sigma_j\}$ se determinan unívocamente y, si A es cuadrada y los σ_j son distintos, los vectores singulares izquierdos y derechos $\{\mathbf{v}_j\}$ y $\{\mathbf{u}_j\}$ se determinan de forma única hasta signos complejos (es decir, factores escalares complejos de módulo 1).

<u>Prueba</u>. Para probar la existencia de la SVD, procedemos por inducción sobre la dimensión de A.

Hagamos $\sigma_1 = ||A||_2$. Como $S = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}||_2 = 1\}$ es compacto, existe un vector $\mathbf{v}_1 \in \mathbb{R}^n$ con $||\mathbf{v}_1||_2 = 1$, y $||\mathbf{u}_1||_2 = \sigma_1$, tales que $\mathbf{u}_1 = A\mathbf{v}_1$.

Extendemos $\mathbf{v}_1$ a una base ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^n$, y $\mathbf{u}_1$ a una base

Descomposición SVD

$\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ ortonormal de $\mathbb{R}^m$. Sean U_1 y V_1 las matrices ortogonales con columnas $\mathbf{u}_i$ y $\mathbf{v}_i$, respectivamente.

Tenemos

$$U^T \Sigma V = S = \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{0} & B \end{pmatrix},$$

donde $\mathbf{w}^T$ es un vector fila, $\mathbf{0}$ es un vector columna y $B \in \mathbb{R}^{(m-1)\times(n-1)}$. Además

$$\Big|\Big| \begin{pmatrix} \sigma_1 & \mathbf{w}^T \\ \mathbf{0} & B \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \Big|\Big|_2 \geq \sigma_1^2 + \mathbf{w}^T\mathbf{w} = (\sigma_1^2 + \mathbf{w}^T\mathbf{w})^{1/2} \Big|\Big| \begin{pmatrix} \sigma_1 \\ \mathbf{w} \end{pmatrix} \Big|\Big|_2,$$

de modo que $||S||_2 \geq (\sigma_1^2 + \mathbf{w}^T\mathbf{w})^{1/2}$.
Como U_1 y V_1 son ortogonales, sabemos que $||S||_2 = ||A||_2 = \sigma_1$, y esto implica que $\mathbf{w} = \mathbf{0}$.

Descomposición SVD

Si $n = 1$ ó $m = 1$, acabó la prueba. De lo contrario, la submatriz B describe la acción de A en el subespacio ortogonal a $\langle \mathbf{v}_1 \rangle$, un espacio de dimensión $n - 1$.

Por la hipótesis de inducción, B admite una descomposición SVD de la forma $B = U_2 \Sigma_2 V_2^T$, y se verifica que

$$A = U_1 \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V_2 \end{pmatrix}^T V_1^T.$$

es una descomposición SVD para A, como queríamos. Esto muestra la existencia.

Para la unicidad, leer el final del capítulo 4, libro de Trefethen. $\square$

Cambio de Bases

La SVD nos permite decir que cada matriz es diagonal, desde que se utilicen las bases adecuadas para los espacios de dominio y rango.
Dado cualquier $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{b}$ puede expandirse en la base de los vectores singulares izquierdos de A (columnas de U), y cualquier $\mathbf{x} \in \mathbb{R}^n$ puede representarse en la base de los vectores singulares derechos de A (columnas de V).

Los vectores de coordenadas para estas expansiones son $\mathbf{b}' = U^T\mathbf{b}$ y $\mathbf{x}' = V^T\mathbf{x}$. De $A = U\Sigma V^T$, la relación $\mathbf{b} = A\mathbf{x}$ se puede expresar en términos de $\mathbf{b}'$ y $\mathbf{x}'$ como

$$\mathbf{b} = A\mathbf{x} \;\Leftrightarrow\; U^T\mathbf{b} = U^TA = U^TU\Sigma V^T\mathbf{x} \;\Leftrightarrow\; \mathbf{b}' = \Sigma\mathbf{x}'.$$

Así, A se reduce a la matriz diagonal Σ cuando el rango se expresa en la base de columnas de U y el dominio en la base de columnas de V.

SVD y Descomposición Espectral

Una matriz cuadrada A de rango completo puede expresarse como una matriz diagonal de autovalores Λ, si el rango y el dominio están representados en una base de autovectores.

Relación entre la SVD y la Descomposición espectral:

Sea $A = U\Sigma V^T$ la SVD de A. Entonces

$$A^TA = (U\Sigma V^T)^T(U\Sigma V) = V\Sigma(U^TU)\Sigma V^T = V\Sigma^2V^T = V\Lambda V^T$$

$\Rightarrow$ las columnas de V son los autovectores de A^TA,
$\Rightarrow$ los valores singulares σ_i son las raíces de los autovalores λ_i de A^TA.

$$AA^T = (U\Sigma V^T)(U\Sigma V)^T = U\Sigma(V^TV)\Sigma U^T = U\Sigma^2U^T = U\Lambda U^T$$

$\Rightarrow$ las columnas de U son los autovectores de AA^T,
$\Rightarrow$ los valores singulares σ_i son las raíces de los autovalores λ_i de AA^T.

Propiedades a partir de la SVD

Sea $A = U\Sigma V^T$ su descomposición en valores singulares.
Las pruebas de los resultados siguientes están en el Capítulo 5 de Trefethen.

Teorema

El rango $\text{rank}(A)$ *es* r*, el número de valores singulares no-nulos.*

Teorema

$\text{Im}(A) = \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$ y $\text{Ker}(A) = \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle$.

Teorema

$||A||_2 = \sigma_1$ y $||A||_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}$.

Propiedades a partir de la SVD

Teorema

Los valores singulares no-nulos de A son las raíces de los autovalores de A^TA o de AA^T. Esto es $\sigma_i = \sqrt{\lambda_i}$, donde $\Lambda = (\lambda_1, \ldots, \lambda_r, \ldots)$ corresponde a la diagonal en la descomposición espectral de A^TA o de AA^T.

Teorema

Si $A = AT$, los valores singulares de A son los valores absolutos $\sigma_i = |\lambda_i|$ de los autovalores de A.

Teorema

Para $A \in \mathbb{R}^{n \times n}$, vale $\det(A) = \prod_{i=1}^{n} \sigma_i$.

Teorema (Eckart-Young)

Sea $A \in \mathbb{R}^{n\times d}$, $n \geq d$, una matriz cuya descomposición SVD está dada por

$$A = USV^T = \sum_{i=1}^{d} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Entonces, la matriz $\widehat{A}_r$ de rango r, $1 \leq r \leq d$, que mejor aproxima A en el sentido de minimizar

$$\min_{rank\ \widehat{A}_r \leq r} ||A - \widehat{A}_r||_F^2$$

se obtiene de truncar la descomposición en valores dingulares de A:

$$\widehat{A}_r = U_r S_r V_r^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

Teorema (Eckart-Young)

donde

$$U_r = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_r],\ \ S_r = diag(\sigma_1, \sigma_2 \ldots, \sigma_r),\ \ V_r = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_r].$$

En ese caso, el error de aproximación está dado por

$$||A - \widehat{A}_r||_F^2 = \sum_{i=r+1}^{d} \lambda_i,$$

o

$$||A - \widehat{A}_r||_2^2 = \lambda_{r+1}.$$

Descomposición de Schur

Otra factorización matricial, en realidad la que es más útil en análisis, es la llamada factoración de Schur.

Definición

*Una **factoración de Schur** de una matriz $A \in \mathbb{R}^{n\times n}$ es una factorización de la forma*

$$A = QTQ^T,$$

donde Q es ortogonal (unitaria) y T es triangular superior.

Observe que A y T son semejantes. De manera similar, los autovalores de A necesariamente aparecen en la diagonal de T.

Teorema

Toda matriz cuadrada A posee una factoración de Schur.

Descomposición de Schur

Prueba: Por inducción sobre la dimensión n de A.

El caso $n = 1$ es directo, ya que $A = [a] = [1][a][1]^T$, así que suponga que $n \geq 2$. Sea $\mathbf{x} \in \mathbb{R}^n$ cualquier autovector de A, con autovalor correspondiente λ.
Tomamos $\mathbf{u}_1 = \frac{\mathbf{x}}{||\mathbf{x}||_2}$ normalizado, y lo hacemos la primer columna de la matriz U.
Entonces, al igual que en la descomposición espectral, vale

$$U^TAU = \begin{pmatrix} \lambda & B \\ \mathbf{0} & C \end{pmatrix},$$

con $B \in \mathbb{R}^{n-1}$, $C \in \mathbb{R}^{(n-1)\times(n-1)}$.
De la hipótesis inductiva, existe una factoración de Schur VTV^T para C. Escribimos entonces

Descomposición de Schur

$$Q = U \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V \end{pmatrix}.$$

Entonces, Q es una matriz ortogonal (unitaria), pues $QQ^T = U \left(\begin{smallmatrix} 1 & 0 \\ 0 & VV^T \end{smallmatrix}\right) U^T = UU^T = I$, y tenemos

$$Q^TAQ = \begin{pmatrix} 1 & 0 \\ 0 & V^T \end{pmatrix} U^TAU \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} \lambda & BV \\ \mathbf{0} & T \end{pmatrix} \Rightarrow A = Q \begin{pmatrix} \lambda & BV \\ \mathbf{0} & T \end{pmatrix} Q^T. \square$$

En resumen:

- Una diagonalización $A = X\lambda X^{-1}$ existe, si y sólo si, A es no defectuosa.
- Una diagonalización unitaria existe, si y sólo si, A es **normal** (esto es, $AA^* = A^*A$).
- Una descomposición SVD $A = U\Sigma V^T$ siempre existe.
- Una descomposición de Schur $A = QTQ^T$ siempre existe.