SVD

Juan Pablo Pinasco (jpinasco@dm.uba.ar//gmail.com)

Departamento de Matemática e IMAS,
FCEyN, UBA - CONICET

2020

Hoy:

- SVD

Próxima:

- Ortonormalización: 113-120

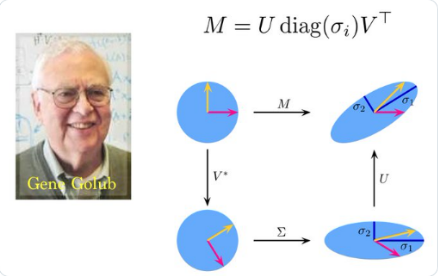
SVD= Singular Values Decomposition = Descomposición en valores singulares

Twittear

Gabriel Peyré
@gabrielpeyre

The SVD decomposes the action of a matrix into rotations and scalings along the axes.
en.wikipedia.org/wiki/Singular_...

Traducir Tweet



3:00 a. m. · 20 feb. 2020 · TweetDeck

95 Retweets **513** Me gusta

1.- SVD

Sagredo -Vamos a generalizar la idea de autovalor.

Simplicio -¿Para qué, si con números complejos siempre tenés autovalores?

Sagredo -Si, el tema es que no toda matriz es diagonalizable (autovalores múltiples y menos autovectores).

Simplicio -Pero hay forma de Jordan!

Sagredo -No toda matriz tiene forma de Jordan...

Simplicio -Pero en Mate 2 nos dijeron...

Sagredo -...por ejemplo si no es cuadrada.

Simplicio Pero... pero... ¿para qué sirve?

Salvati -No se, pero está en el programa.

Los autovalores anda bien si uno quiere potencias de A, o cómo actúa.

- Los autovalores se usan para ver cómo vibran estructuras (edificios, puentes, instrumentos musicales).
- Iterando A^k, los autovalores más grandes dominan la dinámica y se ve el efecto de sus autovectores.
- Es la herramienta que usa Google para rankear las páginas web y elegir qué mostrarte en el buscador.
- Buscás X, ¿en qué páginas está? (y rankearlas por relevancia)

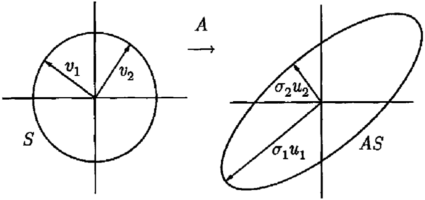
SVD funciona mejor cuando uno necesita conocer la matriz en sí (y no la tiene, o es demasiado grande).

- Descompone la matriz con info directa del núcleo, la imagen, y ciertas direcciones privilegiadas.
- Te da bases ortogonales en $\mathbb{R}^n$, y en $\mathbb{R}^m$, tales que $Av_j = \sigma_j u_j$.
- Te permite invertir (comillas, comillas) cosas no inversibles.
- Es la herramienta que usa Netflix para sugerirte películas o series.
- Te gustó la película X, ¿qué otras películas te gustarán? (y en qué orden)

`http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.241&rep=rep1&type=pdf`

4.- Idea geométrica

Sea $A \in \mathbb{R}^{m\times n}$ con $m \geq n$.
Aplicando A a la esfera unitaria de $\mathbb{R}^n$, la manda a un elipsoide en $\mathbb{R}^m$.

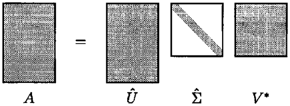


Sean u_j las direcciones del j-ésimo semieje principal del elipsoide; σ_j sus longitudes; y v_j tal que $Av_j = \sigma_j u_j$ (las preimágenes).

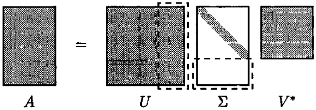
Podemos escribir esto como

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

Tenemos $AV = \hat{U}\hat{\Sigma}$, y resulta que V es unitaria, así que podemos escribir $A = \hat{U}\hat{\Sigma}V^t$. Esta es la descomposición reducida:



Completando a una base, y extendiendo Σ con ceros, tenemos la descomposición SVD, $A = U\Sigma V^t$:



Como $m \geq n$, la imagen seguro que no cubre todo, y por eso se agregan vectores. Se los elige ortogonales con los anteriores.

Tenemos $AV = \hat{U}\hat{\Sigma}$, y resulta que V es unitaria, así que...

¿Por qué V es unitaria?

Porque sé que los semiejes de un elipsoide son ortogonales, me estoy confiando demasiado en la intuición geométrica.

Sería conveniente dar una definición formal, y demostrar todo.

Def: Sea $A \in \mathbb{R}^{m\times n}$ con $m \geq n$. A tiene una descomposición en valores singulares si existen matrices unitarias $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$, y una matriz diagonal $\Sigma \in \mathbb{R}^{m\times n}$, tales que $A = U\Sigma V^t$.

Teor: Sea $A \in \mathbb{R}^{m\times n}$ con $m \geq n$. Entonces, existe una descomposición en valores singulares. Es decir, existen matrices unitarias $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$, y una matriz diagonal $\Sigma \in \mathbb{R}^{m\times n}$ tal que $A = U\Sigma V^t$.

Dem: Sea A^tA, es una matriz simétrica y semidefinida positiva,

$$\langle w, A^tAw\rangle = w^tA^tAw = (Aw)^t \cdot Aw = \|Aw\|^2 \geq 0.$$

Entonces, tiene n autovalores reales positivos $\{\sigma_i^2\}_{i=1:n}$, y una base ortonormal de autovectores asociados $\{v_i\}_{i=1:n}$.

Sean $\{u_i\}_{i=1:r}$, tal que $u_i = \sigma_i^{-1}Av_i$ si $\sigma_i \neq 0$. Estos vectores son l.i., porque si una combinación lineal diera cero, llegaríamos a un absurdo, y también son ortogonales

Meditad con ahínco, hasta hallar en qué consiste (...) este razonamiento. Y cuando os hiervan los sesos, avisad. Antonio Machado.

$$\langle u_i, u_j\rangle = \langle \sigma_i^{-1}Av_i, \sigma_j^{-1}Av_j\rangle = \sigma_i^{-1}\sigma_j^{-1}\langle v_i, A^tAv_j\rangle = \sigma_i^{-1}\sigma_j\langle v_i, v_j\rangle$$

Agreguemos vectores hasta completar una base ortonormal de $\mathbb{R}^n$.

Tenemos :

- $\{\sigma_i^2\}_{i=1:n}$, $\{v_i\}_{i=1:n}$ para A^tA.
- $\{u_i\}_{i=1:n}$, tal que $u_i = \sigma_i^{-1} A v_i$ si $\sigma_i \neq 0$.

Llegamos a la situación del principio, pero ahora sabemos que V, U son matrices unitarias, ya que al multiplicar por su transpuesta tenemos la identidad;

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \left[\begin{array}{c|c|c|c} v_1 & v_2 & \cdots & v_n \end{array}\right] = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \cdots & u_n \end{array}\right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

En definitiva,

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

y el teorema queda demostrado.

Se descompone $A = U\Sigma V^*$ con Σ una diagonal (posiblemente trunca),

Ejemplos de Σ:

$$\Sigma_A = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix} \qquad \Sigma_B = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}$$

Si $m < n$ podemos trabajar con A^t: $A^t = \bar{U}\bar{\Sigma}\bar{V}^t$ y

$$A = \bar{V}\Sigma^t\bar{U}^t$$

Nucleo e imagen se tienen muy rápido de las matrices U y V:

$$\begin{aligned} A &= U\Sigma V^t \\ AV &= U\Sigma V^t V \\ AV &= U\Sigma \end{aligned}$$

Mirando los vectores columnas v_j, u_j de V y U,

$$Av_j = \sigma_j u_j$$

.

- Las columnas de U nos dicen la imagen de A.
- Las columnas de V con $\sigma_j = 0$ son el núcleo de A.
- Los valores más grandes de $|\sigma|$ implican mayores cambios al aplicar A.
- Podemos reducir dimensiones en un problema haciendo $\sigma_j = 0$ si $\sigma_j \leq \epsilon$.
- Por ejemplo, para comprimir imágenes y otros datos que se pueden guardar como valores de una matriz. Tachando algunos σ_j pequeños, esperamos preservar la información principal. Había un ejercicio en la práctica de procesar una foto, pero eran otras épocas (2019, 2018,...)

Sea $A = U\Sigma V^t$, y definamos

$$\Sigma^\dagger = diag(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0).$$

Def: La pseudoinversa de A es

$$A^\dagger = V\Sigma^\dagger U^t.$$

Ejercicio: verificar que se cumplen las propiedades de Penrose, y $X = A^\dagger$ satisface:

i.- $XAX = X$
ii.- $AXA = A$
iii.- $(AX)^t = AX$
iv.- $(XA)^t = XA$

(un teoremita dice que existe una única X que cumple, así que la pseudoinversa es única).