

SVD en la era digital

Dr. Humberto Madrid de la Vega

Universidad Autónoma de Coahuila. México

IV Encuentro Cuba–México de Métodos Numéricos y Optimización.
Enero 2015

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$
- U ortogonal $m \times r$

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$
- U ortogonal $m \times r$
- V ortogonal $n \times r$

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$
- U ortogonal $m \times r$
- V ortogonal $n \times r$
- Σ diagonal $r \times r$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, r es el rango de A .

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$
- U ortogonal $m \times r$
- V ortogonal $n \times r$
- Σ diagonal $r \times r$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, r es el rango de A .

- Las columnas de U forman una base ortonormal para el espacio columna de A

Factorización SVD (Singular Value Decomposition)

- $A = U\Sigma V^T$ A $m \times n$
- U ortogonal $m \times r$
- V ortogonal $n \times r$
- Σ diagonal $r \times r$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, r es el rango de A .

- Las columnas de U forman una base ortonormal para el espacio columna de A
- Las columnas de V forman una base ortonormal para el espacio renglón de A

SVD truncada

A se puede reescribir como

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \quad (1)$$

SVD truncada

A se puede reescribir como

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \quad (1)$$

$$A = U\Sigma V^T = \underbrace{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T}_{A_k} + \underbrace{\cdots + \sigma_r u_r v_r^T}_{\quad}$$

$$A \approx A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T = U_k \Sigma_k V_k^T \quad (2)$$

donde $k < r$, U_k , $m \times k$, Σ_k , $k \times k$, V_k , $n \times k$.

SVD truncada

A se puede reescribir como

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \quad (1)$$

$$A = U\Sigma V^T = \underbrace{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T}_{A_k} + \underbrace{\cdots + \sigma_r u_r v_r^T}_{\quad}$$

$$A \approx A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T = U_k \Sigma_k V_k^T \quad (2)$$

donde $k < r$, U_k , $m \times k$, Σ_k , $k \times k$, V_k , $n \times k$.

A_k tiene rango k y es la mejor aproximación a A de rango k en la norma espectral (norma 2) y en la norma Frobenius

SVD is the Rolls–Royce and the Swiss Army Knife of Numerical Linear Algebra

Dianne O'Leary



- U y V son matrices densas
- El cálculo de la SVD es caro
- No es sencillo actualizar la SVD si A cambia

¿Cómo se puede usar la SVD para matrices enormes?

amazon Prime

Humberto's Amazon.com Today's Deals Gift Cards Sell Help

14 DAYS LEFT Sponsored By Hershey's **Halloween Shop** > Shop now

Shop by Department ▾ Search All ▾ Go

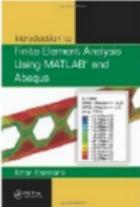
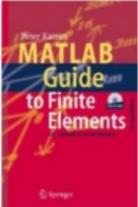
Hello, Humberto **Your Account** Your Prime ▾ Cart ▾ Wish List ▾

Your Amazon.com Your Browsing History **Recommended For You** Improve Your Recommendations Your Profile Learn More

Your Amazon.com

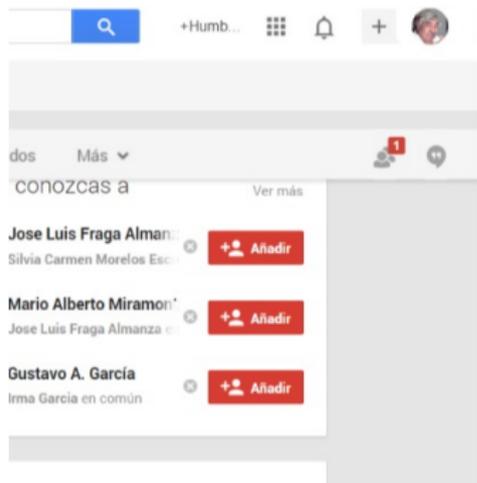
Books

Page 4 of 20 (Start over)

				
<p>Numerical Methods for ... > R. W. Hamming ★★★★★ (25) \$24.95 \$16.02 Why recommended?</p>	<p>Introduction to Finite ... > Amar Khennane ★★★★★ (1) \$129.95 \$123.45 Why recommended?</p>	<p>Computational Fluid ... > John David Anderson ★★★★★ (14) \$228.50 Why recommended?</p>	<p>Fourier Series (Dover ... > Georgil Pavlovic... ★★★★★ (24) \$17.95 \$3.99 Why recommended?</p>	<p>MATLAB Guide to ... > Peter Issa Kattan ★★★★★ (1) \$129.00 \$103.20 Why recommended?</p>

> See all recommendations in Books

Facebook

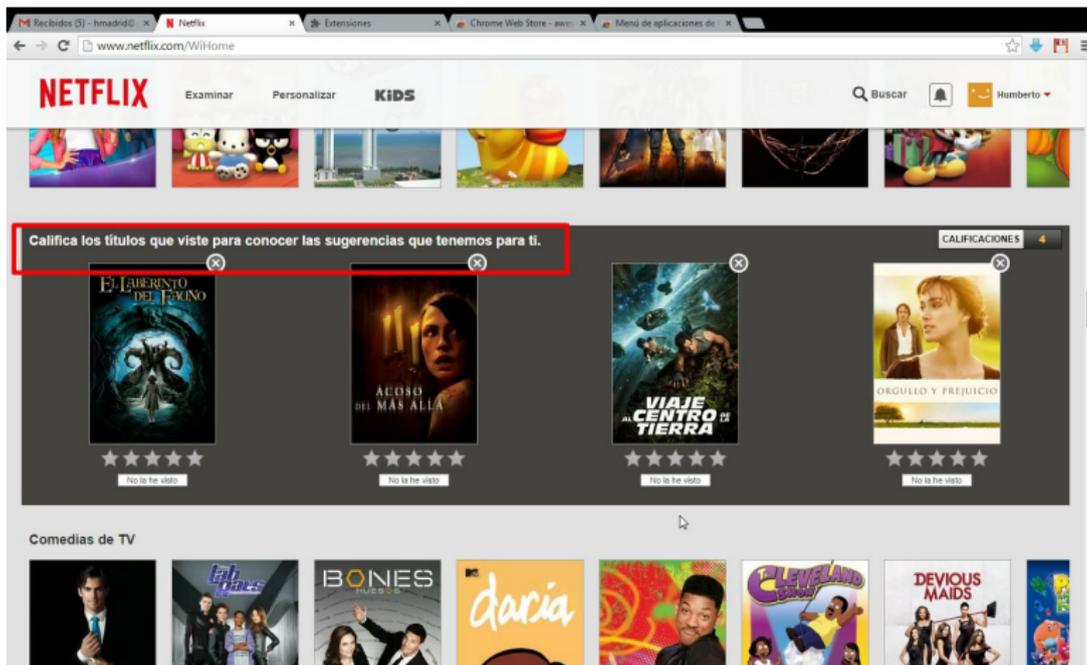


The screenshot shows the Spotify desktop application interface. At the top, there is a search bar with the text "Buscar" and a navigation menu with options: Archivo, Editar, Vista, Reproducción, Ayuda. The main content area is divided into sections: INICIO, ÉXITOS, GÉNEROS Y MOMENTOS, NOVEDADES, NOTICIAS, and **DESCUBRIR**. The **DESCUBRIR** section features a grid of music recommendations with the following content:

- Top Left:** "¿Te gustan Los Bunkers y Toy Selectah? Échale un vistazo a **Panteon Rococo**." Below is an album cover for Panteon Rococo.
- Top Middle:** "Hay gente que escucha a **Los Bunkers** que también escucha a **Fobia**." Below is an album cover for Fobia.
- Top Right:** "Has escuchado a **Los Bunkers**. Aquí tienes un álbum que quizás te guste." Below is an album cover for Solists by León Larregui.
- Bottom Left:** "Si te gusta **Calle 13**, a ver qué te parece **Bunbury**." Below is an album cover for Bunbury.
- Bottom Middle:** "Si te gusta **Bomba Estéreo**, a ver qué te parece **Carla Morrison**." Below is an album cover for Carla Morrison.
- Bottom Right:** "¿Te gustan **Bomba Estéreo** y **Toy Selectah**? Échale un vistazo a **Los Pericos**." Below is an album cover for Los Pericos.

The left sidebar contains navigation options under "INICIO" (Explorar, Actividad, Radio, Seguir, Listas de éxitos, Mensajes, Cola de reproducción, Dispositivos, Buscador de Aplicaciones, musixmatch) and "TU MÚSICA" (Canciones, Álbumes, Artistas, Archivos locales). A "Nueva playlist" button is also present. At the bottom, there is a playback control bar showing a progress bar at 1:38 and a volume icon.

On the right side, there is a "A quién seguir" section with a search bar and a list of artists to follow, including Peter Rdz (26 SEGUIDORES), Ruth Ale Palomino (ha escuchado), Organdi (Jorge Drexler), Fer De León (ha escuchado), Dream of Flying (Plano Solo) (Classical New Age...), Ruth Ale Palomino (ha escuchado), Bolivia (Jorge Drexler), and Fer De León (ha escuchado), Smile Like You Mean It (The Killers).



Premio Netflix

En 2006 Netflix abrió un concurso para premiar con un millón de dólares un algoritmo que mejorara su sistema de recomendación en al menos un 10%. El concurso terminó en 2009.

- Datos de entrenamiento: 100,480,507 valoraciones
- 480,189 usuarios
- 17,770 películas.
- Clasificación [usuario, película, fecha de valoración, valoración].
- Los campos de usuario y de la película son identificadores enteros, valoración de 1 a 5 estrellas.

Sistemas de recomendación

- 1 Usan información (explícita o implícita) sobre las preferencias de los usuarios
- 2 Usan un método para determinar si un artículo es interesante para un usuario

Normalmente, la información de los usuarios incluye información externa, tales como

- Perfiles de los usuarios
- Historiales de compra
- Evaluaciones de productos
- Otras

Recomendación colaborativa

Recomendación colaborativa o filtrado colaborativo (FC).

- Es el enfoque más usado
- Se basa en “la sabiduría de la gente”
- Usa valoración (explícita o implícita) de productos por parte de usuarios
- Hace recomendaciones de productos a nuevos usuarios

Recomendación basada en usuarios

Para un usuario nuevo,

- Encuentra otros usuarios con gustos parecidos
- Con esa información hace recomendaciones al nuevo usuario

Un ejemplo

Un ejemplo típico: recomendación de películas. El escenario es básicamente el siguiente

- Un conjunto de usuarios ha calificado inicialmente algún grupo de películas (por ejemplo, en la escala de 1 a 5) que ya han visto.

Un ejemplo

Un ejemplo típico: recomendación de películas. El escenario es básicamente el siguiente

- Un conjunto de usuarios ha calificado inicialmente algún grupo de películas (por ejemplo, en la escala de 1 a 5) que ya han visto.
- El sistema de recomendación utiliza estas clasificaciones conocidas para predecir las valoraciones que cada usuario daría a esas películas no calificadas por el nuevo usuario.

Ejemplito

4 personas, 6 películas con calificaciones del 1 al 5, el 0 significa que no ha visto la película. Carlos, también ha calificado algunas películas.

	Karen	Juan	Pedro	Luis	Carlos
p_1	5	5	0	5	5
p_2	5	0	3	4	5
p_3	3	4	0	3	0
p_4	0	0	5	3	0
p_5	5	4	4	5	0
p_6	5	4	5	5	5

Problema: basados en la primera tabla, hacer recomendaciones a Carlos.

La idea principal es tratar de determinar quiénes son los que tienen gustos *parecidos* a los de Carlos.

- Hay que usar un criterio de cercanía.

La idea principal es tratar de determinar quiénes son los que tienen gustos *parecidos* a los de Carlos.

- Hay que usar un criterio de cercanía.
- La distancia euclidiana no parece ser una buena opción.

La idea principal es tratar de determinar quiénes son los que tienen gustos *parecidos* a los de Carlos.

- Hay que usar un criterio de cercanía.
- La distancia euclidiana no parece ser una buena opción.
- Una alternativa es usar el coseno del ángulo que forman dos vectores.

La idea principal es tratar de determinar quiénes son los que tienen gustos *parecidos* a los de Carlos.

- Hay que usar un criterio de cercanía.
- La distancia euclidiana no parece ser una buena opción.
- Una alternativa es usar el coseno del ángulo que forman dos vectores.

Esta es un criterio usado por varios motores de búsqueda: Google, Yahoo, MSN, Bing, etc.

Usando este criterio obtenemos

	Karen	Juan	Pedro	Luis
$\cos(\theta)$	0.8295	0.6082	0.5333	0.7742

	Karen(1)	Luis(2)	Carlos
p_1	5	5	5
p_2	5	4	5
p_3	3	3	0
p_4	0	3	0
p_5	5	5	0
p_6	5	5	5

Usamos las valoraciones de Karen y Luis para hacer recomendaciones a Carlos sobre las películas p_3 , p_4 y p_5

k vecinos más cercanos

La técnica utilizada es la de k vecinos más cercanos

- Para fines prácticos, pensemos en que la matriz de prueba de Netflix es N de tamaño $17,000 \times 500,000$
- Cada vez que se hace una nueva recomendación la matriz se modifica y las recomendaciones pueden cambiar mucho.
- Hay que recomputar todo con mucha frecuencia.
- Si tenemos una base de datos muy grande este procedimiento es muy costoso.

- Podemos intentar *comprimir* los datos.

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj
 - Imágenes: jpg, gif, tif

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj
 - Imágenes: jpg, gif, tif
 - Video: avi, mpeg, mov, flv

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj
 - Imágenes: jpg, gif, tif
 - Video: avi, mpeg, mov, flv
 - Audio: mp3, aif, au, ogg

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj
 - Imágenes: jpg, gif, tif
 - Video: avi, mpeg, mov, flv
 - Audio: mp3, aif, au, ogg
- Se puede usar una técnica de *reducción de dimensión*.

- Podemos intentar *comprimir* los datos.
- Ejemplos de compresión de datos
 - Archivos: zip, rar, arj
 - Imágenes: jpg, gif, tif
 - Video: avi, mpeg, mov, flv
 - Audio: mp3, aif, au, ogg
- Se puede usar una técnica de *reducción de dimensión*.
- Descomposición en Valores Singulares (SVD)

Reducción de dimensión

Vectores coordenados

Sea S un subespacio de \mathbb{R}^n de dimensión k . Sea x_1, x_2, \dots, x_k una base de S . Sea $y \in S$, entonces

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k = X\alpha$$

donde X es la matriz cuyas columnas son las x_j y α es un vector $k \times 1$, cuyos elementos son las α_j . Se dice que α es el vector coordenado de y en la base de S . **Entonces aunque $y \in \mathbb{R}^n$, como elemento del subespacio S , lo podemos representar con un vector α de k componentes**

Proyección sobre subespacios

Proyección ortogonal

Sea S un subespacio de \mathbb{R}^n de dimensión k . Sea x_1, x_2, \dots, x_k **una base ortonormal de S** . Sea $y \notin S$, entonces z , la proyección ortogonal de y sobre S , es $z = XX^T y = X(X^T y)$, por lo tanto el vector coordenado de z en la base X es $X^T y$

Regresando a las recomendaciones

$A = [a_1, a_2, \dots, a_n]$ matriz $m \times n$ (por ejemplo películas–usuarios) y un nuevo usuario q . Queremos hacerle a p recomendaciones sobre películas que no ha visto.

La idea principal es proyectar las columnas de A y a q sobre el subespacio generado por las columnas de U_k

$$A = U\Sigma V^T = \underbrace{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T}_{U_k} + \underbrace{\dots + \sigma_r u_r v_r^T}_{U_k^\perp}$$

$$a_j = Ae_j \approx A_k e_j = U_k (\Sigma_k V_k^T e_j)$$

El vector coordenado de la proyección de a_j sobre U_k es $(\sum_k V_k^T) e_j$
Este último vector es la columna j de $\sum_k V_k^T$, entonces las columnas de esta matriz son los vectores coordenados de las columnas de A , las cuales representan a los usuarios ahora en \mathfrak{R}^k (recordemos que $\sum_k V_k^T$ es de orden $k \times n$)

El vector coordenado de la proyección de a_j sobre U_k es $(\sum_k V_k^T) e_j$.
Este último vector es la columna j de $\sum_k V_k^T$, entonces las columnas de esta matriz son los vectores coordenados de las columnas de A , las cuales representan a los usuarios ahora en \mathfrak{R}^k (recordemos que $\sum_k V_k^T$ es de orden $k \times n$)

Ahora proyectamos q sobre U_k , esta proyección es $U_k U_k^T q$, así que su vector coordenado es $p = U_k^T q$.

El vector coordenado de la proyección de a_j sobre U_k es $(\sum_k V_k^T) e_j$.
Este último vector es la columna j de $\sum_k V_k^T$, entonces las columnas de esta matriz son los vectores coordenados de las columnas de A , las cuales representan a los usuarios ahora en \mathfrak{R}^k (recordemos que $\sum_k V_k^T$ es de orden $k \times n$)

Ahora proyectamos q sobre U_k , esta proyección es $U_k U_k^T q$, así que su vector coordenado es $p = U_k^T q$.

Ahora buscamos entre las columnas de $\sum_k V_k^T$ aquellas que estén más cercanas a p .

De nuevo el ejemplito

	Karen	Juan	Pedro	Luis	Carlos
p_1	5	5	0	5	5
p_2	5	0	3	4	5
p_3	3	4	0	3	0
p_4	0	0	5	3	0
p_5	5	4	4	5	0
p_6	5	4	5	5	5

Así que

$$A = \begin{bmatrix} 5 & 5 & 0 & 5 \\ 5 & 0 & 3 & 4 \\ 3 & 4 & 0 & 3 \\ 0 & 0 & 5 & 3 \\ 5 & 4 & 4 & 5 \\ 5 & 4 & 5 & 5 \end{bmatrix} \quad q = \begin{bmatrix} 5.0 \\ 5.0 \\ 0 \\ 0 \\ 0 \\ 5.0 \end{bmatrix}$$

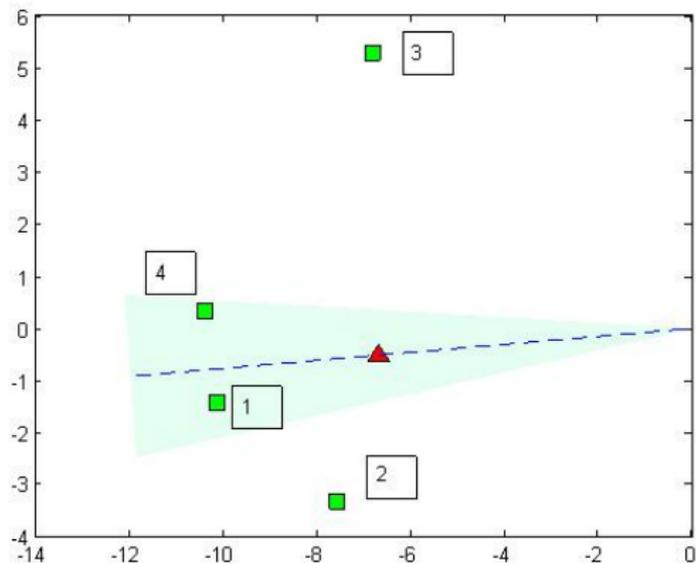
Tomemos $k = 2$

Tomemos $k = 2$

$$U_2 = \begin{bmatrix} -0.4472 & -0.5373 \\ -0.3586 & 0.2461 \\ -0.2925 & -0.4033 \\ -0.2078 & 0.67 \\ -0.5099 & 0.0597 \\ -0.5316 & 0.1887 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 17.71 & 0 \\ 0 & 6.392 \end{bmatrix}$$

$$V_2^T = \begin{bmatrix} -0.571 & -0.4275 & -0.3846 & -0.5859 \\ -0.2228 & -0.5172 & 0.8246 & 0.0532 \end{bmatrix}$$

$$\Sigma_2 V_2^T = \begin{bmatrix} -10.11 & -7.572 & -6.813 & -10.38 \\ -1.424 & -3.306 & 5.271 & 0.34 \end{bmatrix}$$



Los más cercanos a Carlos (p), son Karen (1) y Luis (4)

Podemos también hacer recomendaciones basados en la popularidad de las películas,

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

- SVD truncada compacta la información desechando información superflua.

- SVD truncada compacta la información desechando información superflua.
- Conviene trabajar mejor con A_k , $k < r$

- SVD truncada compacta la información desechando información superflua.
- Conviene trabajar mejor con A_k , $k < r$
- Revela información instrínseca (latente)

- SVD truncada compacta la información desechando información superflua.
- Conviene trabajar mejor con A_k , $k < r$
- Revela información instrínseca (latente)
- Usada en LSA (Latent Semantic Analysis)

- SVD truncada compacta la información desechando información superflua.
- Conviene trabajar mejor con A_k , $k < r$
- Revela información instrínseca (latente)
- Usada en LSA (Latent Semantic Analysis)
- Motores de búsqueda: polisemia, sinonimia.

- Memoria: Matrices muy grandes
- Cómputo muy pesado
- Muchísimos ceros (matrices sparse)
- Actualización de la matriz

Actualización vs Folding-in

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

Si le agregamos una columna a A su SVD cambia completamente. Actualizar la SVD es calcular de nuevo U , Σ y V .

Folding-in

Pensemos en $U^T A = \Sigma V^T$. Si agregamos una columna p a A

$$U^T [A p] = [U^T A \quad U^T p] = [\Sigma V^T, U^T p] = \Sigma \tilde{V}^T$$

donde

$$\tilde{V} = \begin{bmatrix} V \\ p^T U \Sigma^{-1} \end{bmatrix}^T$$

entonces

$$[A p] = U \Sigma \tilde{V}$$

Incremental SVD

- Se puede usar una primera submatriz y su SVD

Incremental SVD

- Se puede usar una primera submatriz y su SVD
- Con ello se pueden hacer recomendaciones a todo un grupo de usuarios usando folding-in

Incremental SVD

- Se puede usar una primera submatriz y su SVD
- Con ello se pueden hacer recomendaciones a todo un grupo de usuarios usando folding-in
- En realidad la situación es más complicada

Incremental SVD

- Se puede usar una primera submatriz y su SVD
- Con ello se pueden hacer recomendaciones a todo un grupo de usuarios usando folding-in
- En realidad la situación es más complicada
- Hay otros algoritmos Incremental SVD

Incremental SVD

- Se puede usar una primera submatriz y su SVD
- Con ello se pueden hacer recomendaciones a todo un grupo de usuarios usando folding-in
- En realidad la situación es más complicada
- Hay otros algoritmos Incremental SVD
- Área activa de investigación

- Koren, Y. Matrix Factorization Techniques for Recommender Systems. Computer. 2009.
- Sarwar et. al. Incremental Singular Value Decomposition Algorithm for Highly Scalable
- Application of Dimensionality Reduction in Recommender System-A Case Study. Recommender Systems
- Percy, M. Collaborative Filtering for Netflix