El Álgebra Lineal en el entorno personal de las peticiones a Internet (recomendaciones a: música, películas, libros, compras, etc.)

Dr. Humberto Madrid de la Vega
hmadrid@gmail.com

Dra. Marisol Flores Garrido
maria.flores@gmail.com

México

Resumen

Cuando un sitio web sugiere un libro, una película, un artículo o a personas conocidas en una red social, está haciendo uso de un sistema de recomendación que realiza sugerencias basándose en el historial y los patrones de uso y consumo del usuario. En este trabajo mostramos la forma en que los sistemas de recomendación utilizan conceptos y herramientas de Álgebra Lineal. Reconocer esta relación hace posible el uso de dichos sistemas como una herramienta didáctica para la enseñanza del Álgebra Lineal, pues motiva un aprendizaje significativo y enriquece el material del curso al proveer ideas concretas sobre las cuales los estudiantes pueden construir un entendimiento de conceptos abstractos, como espacios vectoriales, distancias, bases y factorizaciones matriciales, entre otros. Adicionalmente puede usarse como un ejemplo de modelación matemática. Los datos se representan como vectores y matrices. Para conjuntos de datos de tamaño modesto se puede usar herramienta relativamente sencilla y para conjuntos de datos mayores se ve la necesidad de reducir la dimensión del problema llevando esto a la necesidad de introducir herramienta más sofisticada.

Palabras clave. Vectores, matrices, producto escalar, factorizaciones matriciales, matrices ortogonales, descomposición SVD.

El Cálculo y su Enseñanza. Volumen 9, Jul - Dic 2017, Cinvestav-IPN, Ciudad de México, pp. **64-76**
Fecha de recepción: 08-08-2017
Fecha de aceptación: 01-11-2017

1. Introducción

"Considera esta paradoja: por un lado, las matemáticas son el hilo que entreteje nuestra vida diaria. Cada vez que compramos algo en línea, enviamos un mensaje de texto, hacemos una búsqueda en Internet o utilizamos un GPS, entran en acción todo tipo de fórmulas matemáticas y algoritmos. Por otro lado, (...) la mayoría de las personas piensa, con una mezcla de orgullo y desafío, que las matemáticas son pura tortura, una pesadilla"

Desde ciencias biológicas hasta ingeniería o el análisis de redes de todo tipo, la comunidad científica cuenta con cantidades impresionantes de datos. Álgebra Lineal es un lenguaje que permite encontrar estructura, manipular e interpretar la información. A final de cuentas la tecnología digital está basada en el Álgebra Lineal.

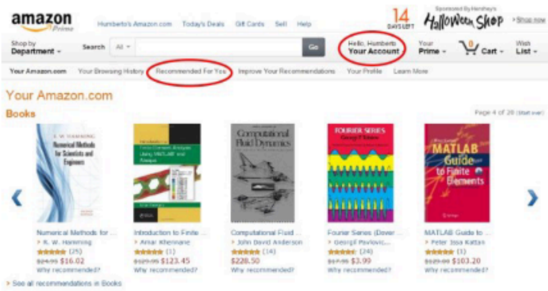
En términos educativos hay una serie de obstáculos. Muchos estudiantes no aprueban - o aprueban sin comprender bien los conceptos (y son incapaces de aplicarlos en situaciones de la vida real); es un curso muy demandante cognitivamente; se introducen muchas ideas, en tiempo limitado, que tienen poca conexión con material estudiado previamente. J. L. Dorier [Dorier 2002] señala que "*Los estudiantes suelen sentir que ellos aterrizan en otro planeta, ellos se sienten abrumados por el gran por el número de definiciones nuevas y la falta de conexión con conocimiento previo. Por otro lado, los profesores suelen sentirse frustrados y desarmados cuando se enfrentan a la incapacidad de sus estudiantes para lidiar con las ideas que consideran que son sencillas*"

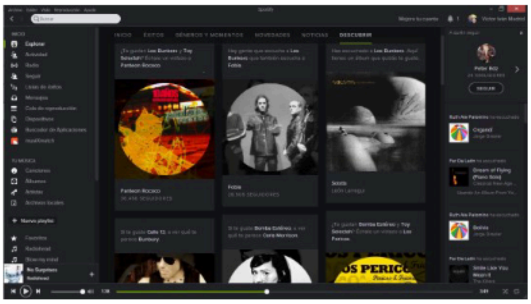
Nuestro objetivo es mostrar el uso del Álgebra Lineal en lo que se denomina **Sistemas de Recomendación**

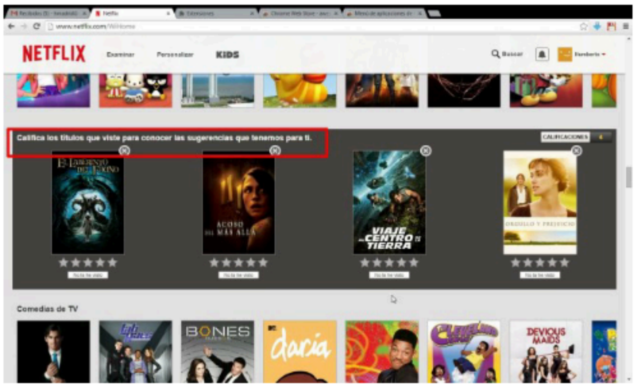
Los sistemas de recomendación representan justamente un intento por ayudarnos a navegar entre opciones. Todos nos hemos topado con ellos en Internet cuando Facebook nos sugiere "personas que tal vez conozcas", cuando Amazon nos cuenta que los usuarios que compraron tal libro también se llevaron tales otros o cuando Netflix sugiere que como nos gustó la película X ahora podríamos ver la Y, que también es un documental biográfico de un músico, por ejemplo. Pero ¿cómo puede una computadora adivinar lo que me va a gustar? Bueno, aunque en gustos se rompen géneros, las personas generalmente seguimos ciertos patrones. Si mi película favorita es *El conjuro*, por ejemplo, es probable que también me haya gustado *El exorcista*. Los sistemas de recomendación tratan de identificar los patrones que seguimos las personas. Así pues, los sistemas de recomendación usan información (explícita o implícita) sobre los usuarios (perfil, historial de compra, evaluación de productos) y tratan, a partir de esa información, de determinar si un artículo es interesante para un usuario en particular.

2.Sistemas de recomendación

Las siguientes figuras muestran algunos ejemplos de recomendaciones que quizá te resulten familiares, pero, en general, este tipo de sistemas ha proliferado en la web en la última década. De alguna forma, los sistemas de recomendación son como el dependiente atento de una tienda, que se ofrece a ayudarnos a encontrar el producto que necesitamos, nos señala opciones, nos hace sugerencias sobre artículos parecidos.







Los sistemas de recomendación tratan de identificar los patrones que seguimos las personas.
Así pues, los sistemas de recomendación usan información (explícita o implícita) sobre los usuarios (perfil, historial de compra, evaluación de productos) y tratan, a partir de esa información, de determinar si un artículo es interesante para un usuario en particular.

Existen tres tipos fundamentales de sistemas de recomendación:

- *Basados en contenido*, que hacen una recomendación basada en las características de los objetos y la similitud entre ellos
- *Filtrado colaborativo*, que se enfoca en la relación entre usuarios y objetos. Las recomendaciones (e implícitamente, la similitud entre objetos) se determinan a partir de la similitud entre los gustos de usuarios.
- *Híbridos*, que combinan de alguna forma los enfoques anteriores

En general, un sistema de recomendación **tiene que** tener al menos dos características importantes. En primer lugar, debe ser rápido, pues no puede tardar horas en buscar recomendaciones para cada usuario. En segundo lugar, tiene que ser eficaz ya que tiene que hacer una buena recomendación si quiere conservar la confianza del usuario en el sitio web que hace la recomendación.

Aquí abordaremos el filtrado colaborativo. Cuando una persona decide, por ejemplo, ir al cine es frecuente que solicite recomendaciones de películas a personas cercanas o busque opiniones sobre ellas en Internet y a final de cuentas se toma más en cuenta a las personas afines en gustos. El filtrado colaborativo intenta modelar esta situación. El término **colaborativo** se refiere a que para tomar una decisión se recurre a la colaboración de otras personas, por ejemplo, y el término **filtrado** se refiere a que discriminamos algunas opiniones.

En resumen, el filtrado colaborativo (FC) se basa en "la sabiduría de la gente". Para hacer recomendaciones a un usuario, usa valoración de productos por parte de otros usuarios, encuentra usuarios con gustos parecidos y con esa información hace recomendaciones al usuario.

¿Cómo puede ayudar el Álgebra Lineal para hacer recomendaciones?

2.1 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.

Digamos que existe un conjunto de 4 usuarios – Karen, Juan, Pedro, Luis– que ya han calificado 6 películas $p_{1,} \dots, p_6$ usando una escala del 1 a 5 (0 significa que el usuario no ha visto la película), y se quiere utilizar esa información para hacerle una recomendación a un usuario nuevo, Carlos.

El primer paso consistiría en sondear un poco los gustos de Carlos, para enterarnos qué calificación le pone a unas cuantas (relativamente pocas) películas. Por ejemplo, digamos que nos cuenta que amó (y calificó con 5) las películas p_1, p_2 y p_6.

Ahora quisiéramos aprovechar la información que tenemos sobre Karen, Juan, Pedro, Luis para "adivinar" qué película le gustará a Carlos. ¿Cómo hacemos esto?
Por ejemplo, supongamos que

- Karen calificó a las películas p_1, p_2, p_3, p_5 y p_6, con los valores 5, 5, 3, 5, 5, respectivamente.
- Juan calificó a p_1, p_3, p_5 y p_6, como 5, 4, 4, 4, respectivamente.
- Pedro calificó p_2, p_4, p_5 y p_6, como 3, 5, 4, 5, respectivamente.
- Luis calificó a p_1, p_2, p_3, p_4 p_5 y p_6 como 5, 4, 3, 3, 5, 5, respectivamente.

Recordemos que Carlos calificó p_1, p_2 y p_6 con 5. ¿Qué película le recomendaríamos a Carlos? ¿a qué usuario(s) se parece en sus gustos?

2.2 Un modelo vectorial

Para hacer el trabajo más fácil, conviene organizar la información en las siguientes tablas

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

Carlos
5
5
0
0
0
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. Organizada la información de esta manera es fácil observar que Karen y Luis son más cercanos a Carlos en gustos cinematográficos. Por supuesto si tuviéramos más información (por ejemplo, más películas y más usuarios) es complicado buscar los usuarios más cercanos por mera observación.

Antes de seguir adelante conviene destacar que básicamente lo que hemos hecho es asociar a cada persona un vector columna que consta de sus calificaciones. Los vectores asociados a Karen, Juan, Pedro y Luis, constituyen las columnas de la matriz películas-usuarios. Esto constituye un modelo vectorial o modelo matricial.

El problema en términos matemáticos es determinar cuáles columnas de la matriz se parecen más al vector asociado al nuevo usuario. Es decir, se reduce a introducir un criterio de cercanía entre vectores.

Lo natural es pensar primeramente en la distancia euclidiana entre vectores. Sin embargo, en este contexto no es la mejor opción. Una razón es que la escala de calificaciones es arbitraria y así por ejemplo no podemos decir una calificación de 5 es mejor que una calificación de 1.

Un criterio usado en el análisis de datos para medir la cercanía entre vectores de datos, es el ángulo entre ellos. Entonces una alternativa es usar el coseno del ángulo que forman dos vectores columna x, y.

$$\cos(\theta) = \frac{x^T y}{\|x\| \|y\|}$$

Usando este criterio obtenemos

	Karen	Juan	**Pedro**	Luis
$\cos(\theta)$	0.8295	0.6082	0.5333	0.7742
θ	33.9	52.5	57.8	39.3

Considerando solamente a Karen y Luis tenemos

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

Carlos
5
5
0
0
0
5

Usamos las valoraciones de Karen y Luis para hacer recomendaciones a Carlos sobre las películas p_3 y p_5. Por ejemplo podemos tomar el promedio de las valoraciones de Karen y Luis, obteniendo 3 para p_3 y 5 para p_5.

Se pueden abordar ejemplos más grandes y no solamente sobre recomendaciones de películas, puede hacerse de libros, de música, etc.

2.3 Algoritmo

Primera parte

- Formar la matriz de datos $A = [a_1, a_2, \ldots, a_n]$ y el vector u correspondiente al nuevo usuario.
- Calcular

$$\cos(\theta_j) = \frac{a_j^T u}{\|a_j\| \|u\|}, \quad j = 1, \ldots, n$$

- Escogemos una k y encontramos los k valores de $\cos(\theta_j)$ más cercanos a 1

Segunda parte

- A cada uno de los items que no haya valorado u se le asigna un valor, por ejemplo el promedio de las valoraciones de los k elementos seleccionados
- O bien un promedio ponderado o algún otro criterio

2.4 Una simplificación

Retomemos la expresión

$$\cos(\theta_j) = \frac{a_j^T u}{\|a\|_j \|u\|}, \quad j = 1, \dots, n$$

Si a_j y u estuvieran normalizados, es decir $\|a_j\| = \|u\| = 1$, entonces

$$\cos(\theta_j) = a_j^T u, \quad j = 1, \dots, n$$

y

$$A^T u = \begin{bmatrix} a_1^T u \\ a_2^T u \\ \vdots \\ a_n^T u \end{bmatrix} = \begin{bmatrix} \cos(\theta_1) \\ \cos(\theta_2) \\ \vdots \\ \cos(\theta_n) \end{bmatrix}$$

Primera parte modificada

- Formar la matriz de datos $A = [a_1, a_2, \dots, a_n]$ y el vector u correspondiente al nuevo usuario.
- Formar la matriz $\tilde{A}$ con las columnas de A normalizadas y $\tilde{u}$ normalizando u
- Calcular $\tilde{A}^T \tilde{u}$ que contiene los valores de $\cos(\theta_j)$
- Escogemos una k y encontramos los k valores de $\cos(\theta_j)$ mÁs cercanos a 1
- El algoritmo es sencillo, fácil de programar
- El tamaño de la matriz A puede ser hasta del orden de cientos
- Se puede aplicar a recomendaciones de películas, canciones, grupos musicales, libros, etc.
- Permite que los mismos estudiantes generen ejemplos y los pongan en práctica

3.Ampliando el horizonte

El 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.

La matriz A es entonces de tamaño $17{,}770 \times 480{,}189$ y cada u es de $17{,}770 \times 1$.

El sentido común nos dice que el algoritmo anterior no es apropiado porque sería muy lento.
El equipo ganador del concurso de Netflix, usó una técnica basada en Álgebra Lineal, concretamente una Factorización matricial llamada SVD (Singular Value Decomposition)

El problema al que nos enfrentamos es el de hacer más económico el proceso, pero sin demérito de los resultados.

3.1 Sobre el producto matriz–vector

La primera parte del algoritmo se puede reducir al cálculo de una matriz por un vector.
¿Podemos realizar este producto con menos operaciones?

Respuesta: Es posible, dependiendo de las características de la matriz.

En general si A es $m \times n$ y x es $n \times 1$, el producto Ax requiere $m \times n$ multiplicaciones.

Consideremos un caso particular

$$A = uv^T, \quad u, m \times 1, \quad v, n \times 1$$

Truco: Usando la asociatividad del producto

$$Ax = (uv^T)x = u(v^T x)$$

Observemos

- $v^T x$ es un número real, un producto escalar, requiere n multiplicaciones
- $u(v^T x)$ es el producto de un escalar por un vector, requiere m multiplicaciones
- $u(v^T x)$ requiere ¡$m + n$ multiplicaciones!

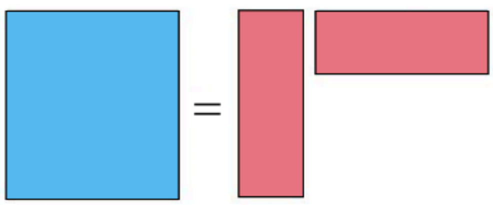
Es decir, el producto Ax requiere $m \times n$ multiplicaciones, en cambio $u(v^T x)$ requiere $m + n$ multiplicaiones. Por ejemplo si $m = n = 100$, entonces $m \times n = 1{,}000{,}000$, pero $m + n = 2{,}000$, lo cual hace una buena diferencia.

Moraleja: El orden de los factores no afecta el producto, pero sí puede afectar el número de operaciones para calcularlo.

3.2 Factorizaciones matriciales y ahorro de operaciones

El ejemplo anterior es el caso más sencillo de una Factorización matricial. Nos interesa una Factorización de la forma

$$A = BC$$



Afirmación. Siempre es posible factorizar una matriz A, $m \times n$ como

$$A = BC$$

donde B es $m \times r$, C es $r \times n$, siendo r el rango de A.

Recordemos que el rango de A es el número de columnas linealmente independientes de A y que $r \leq m, n$.

Si $r \ll m, n$, entonces podemos ahorrar muchas operaciones. Para ver esto, vamos a retomar la expresión

$$\cos(\theta_j) = \frac{a_j^T u}{\|a_j\| \|u\|}, \quad a_j, u \in \mathbb{R}^m$$

Supongamos $A = BC$, $C = [c_1, c_2, \ldots, c_n]$. Entonces $a_j = Bc_j$ y sustituyendo en la expresión anterior tenemos

$$\cos(\theta_j) = \frac{c_j^T (B^T u)}{\|a_j\| \|u\|}, \quad j = 1, \ldots, n$$

Observemos que ahora $c_j, B^T u \in \mathbb{R}^r$, por lo tanto ahorramos operaciones en el numerador. Podemos ahorrar más operaciones si hacemos una suposición adicional. Supongamos que B tiene columnas ortonormales, entonces $\|Bx\| = \|x\|$. En consecuencia

$$\|a_j\| = \|Bc_j\| = \|c_j\|$$

y

$$\cos(\theta_j) = \frac{c_j^T (B^T u)}{\|c_j\| \|u\|}, \quad j = 1, \ldots, n$$

los vectores involucrados, a excepción de u, pertenecen a $\mathbb{R}^r$.

Resumen

Si suponemos que $A = BC$ y

- B es $m \times r$, C es $r \times n$, siendo r el rango de A
- $r \ll m, n$
- B tiene columnas ortonormales

entonces

$$\cos(\theta_j) = \frac{c_j^T (B^T u)}{\|c_j\| \|u\|}, \quad j = 1, \ldots, n$$

los vectores involucrados, a excepción de u, pertenecen a $\mathbb{R}^r$.

Una modificación

En lugar de

$$\cos\left(\theta_j\right) = \frac{c_j^T\left(B^T u\right)}{\left\|c_j\right\|\|u\|}, \quad j = 1, \dots, n$$

tomamos

$$\cos\left(\widehat{\theta}_j\right) = \frac{c_j^T\left(B^T u\right)}{\left\|c_j\right\|\left\|B^T u\right\|}, \quad j = 1, \dots, n$$

Con esta modificación todos los vectores involucrados están en $\mathbb{R}^r$, siendo r el rango de la matriz A. Si $r \ll m, n$ nos ahorramos muchas operaciones.

El proceso anterior es conocido como **reducción de dimensión**, es un caso especial de compresión de datos.

Ejemplos de compresión de datos

- Archivos: zip, rar, arj
- Imágenes: jpg, gif, tif
- Video: avi, mpeg, mov, flv
- Audio: mp3, aif, au, ogg

3.3 ¿Quiénes son B y C?

Existen dos factorizaciones que pueden ser usadas

- Factorización QR: $A = QR$, Q columnas ortonormales, R triangular superior, entonces $B = Q$ y $C = R$
- Descomposición SVD: $A = U\Sigma V^T$, U, V columnas ortonormales, Σ diagonal. Usualmente se toman $B = U$ y $C = \Sigma V^T$

Paquetes como Matlab, Octave, Scilab, Mathematica, etc. tienen implementadas estas factorizaciones.

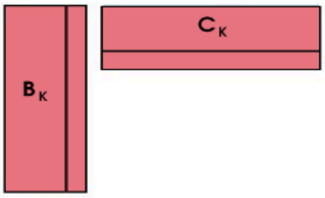
3.4 Más reducción de dimensión

La idea de reducción de dimensión es más general. Lo podemos plantear así. Encontrar B y C tales que

$$A = BC \approx B_k C_k$$

donde

- B es $m \times r$, C es $r \times n$, siendo r el rango de A
- B con columnas ortonormales
- B_k $m \times k$, k columnas de B
- C_k $k \times n$, k filas de C
- $k < r$



3.5 De nuevo el ejemplo

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

Carlos
5
5
0
0
0
5

Así que

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

Haremos uso de la descomposición SVD

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

donde $B = U$ y $C = \Sigma V^T$. B es 6×6 y C es 6×4.

Aproximaremos

$$A \approx B_2 C_2$$

es decir $k = 2$, B_2 es 6×2, consta de las primeras dos columnas de B y C_2 es 2×4, se forma con las primeras dos filas de C.

Luego usaremos

$$\cos(\widehat{\theta}_j) = \frac{c_j^T (B_2^T u)}{\|c_j\| \|B_2^T u\|}, \quad j = 1, \dots, 4$$

Concretamente

$$B_2 = \begin{bmatrix} -0.4472 & -0.5373 \\ -0.3586 & 0.2461 \\ -0.2925 & -0.4033 \\ -0.2078 & 0.67 \\ -0.5099 & 0.0597 \\ -0.5316 & 0.1887 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} -10.11 & -7.572 & -6.813 & -10.38 \\ -1.424 & -3.306 & 5.271 & 0.34 \end{bmatrix}, \quad B_2^T u = \begin{bmatrix} -6.6873 \\ -0.5126 \end{bmatrix}$$

Calculamos los ángulos entre $B_2^T u$ y las columnas de C_2

$$[3.6°, 19.2, 42.1, 6.2°]$$

Los más cercanos a Carlos son de nuevo Karen y Luis.

4. Algo más sobre factorizaciones matriciales.

Afirmación. Siempre es posible factorizar una matriz A, $m \times n$ como

$$A = BC$$

donde B es $m \times r$, C es $r \times n$, siendo r el rango de A.

Para convencernos de ello comenzamos con unas observaciones

Observación 1

$$A = BC = B[c_1, c_2, \dots, c_n] = [Bc_1, Bc_2, \dots, Bc_n]$$

Entonces cada columna de A tiene la forma

$$a_j = Bc_j$$

Observación 2. El producto matriz–vector es una combinación lineal de las columnas de A

$$Ax = x_1a_1 + x_2a_2 + \cdots + x_na_n$$

En nuestro caso

- Cada $a_j = Bc_j$ es combinación lineal de las columnas de B
- Las columnas de B son ortonormales
- Ellas forman una base para el espacio columna de A cuya dimensión es r
- El problema de encontrar B es el de construir una base ortonormal de dicho espacio columna
- Una forma de construir B es mediante el proceso de Gram–Schmidt
- Esto lleva a la factorización QR
- Otra forma de obtener una base ortonormal es con la descomposición SVD

Por lo general las matrices asociadas a bases de datos contienen mucha información redundante y muchos datos se pueden repetir o bien muchos datos muy parecidos entre sí. Para matrices del orden de miles, se puede usar una k modesta, digamos entre 100 y 200.

En resumen, el tema de Sistemas de recomendación se puede abordar

- Mediante un modelo vectorial, asociándole una matriz y un vector
- La búsqueda de los usuarios cercano a un nuevo usuario de puede realizar mediante productos escalares o el producto matriz–vector
- Se puede reducir la dimensión del problema, con el consecuente ahorro de operaciones, mediante factorizaciones matriciales
- También es posible abordar el tema mediante espacios vectoriales, bases ortonormales, proyecciones ortogonales, etc. proporcionando contenido concreto y significativo a esos conceptos y sus propiedades.

5. Referencias

J. L. Dorier. Teaching Linear Algebra at University. ICM 2002, Vol. Ill, 875-884.

G. Harel, Principles of Learning and Teaching Mathematics, With Particular Reference to the Learning and Teaching of Linear Algebra: Old and New Observations, in J-L. Dorier (Ed.), *On the Teaching of Linear Algebra,* Kluwer Academic Publishers, Dordrecht, 2000, 177-189.

G. Strang. Introduction to Linear Algebra. Wellesley - Cambridge Press. Cuarta edición. 2009.