Factorización no negativa de matrices

Una aplicación a los motores de recomendación

Carlos J. Gil Bellosta

cgb@datanalytics.com

Febrero de 2015

¿Son así nuestros súper súper clientes?

Fuente: Rich and Famous http://rf.ro

Datos "diádicos" (¿matrices?)

	T_1	T_2	...	T_m
C_1	0	3	...	1
C_2	2	1	...	0
...	...	...	...	0
C_n	1	0	...	1

Tenemos n clientes (filas) y m (tipos de) productos (columnas).

Objetivo

Entender (¿microsegmentar?) los gustos y preferencias de los clientes.

Lo primero que se te ocurre: SVD

La descomposición en valores singulares (SVD) de X:

$$X = UDV$$

donde:

- D es una matriz diagonal con entradas $d_1 \geq d_2 \geq \cdots \geq 0$.
- Las columnas (filas) de U (V) son ortogonales.

Recomposición del comportamiento de un cliente

El comportamiento de un cliente *es* su fila de X. El comportamiento del cliente i sería

$$X_i = \sum_j u_{ij} d_j v_j.$$

Interpretación:

- Las filas de V, v_j son *preferencias* comunes a todos los clientes.
- Cada cliente asigna un valor a dichas preferencias en función de su fila (la i) en U.
- Los valores d_i son pesos globales de cada *preferencia*.

Reducción de la dimensionalidad

¿Qué si ignoramos aquellas preferencias de menor peso?

Interpretación de las preferencias

¡Buena suerte con los signos negativos!

Pros, cons y un canje razonable

- Nos gusta la descomposición $X = UDV$.
- Nos gusta la posibilidad de reducir la dimensionalidad de la matriz original (aun perdiendo algo de precisión).
- Nos gustaría poder canjear la ortogonalidad por otra propiedad que facilitase la interpretación.

Un canje razonable

Perder la ortogonalidad a cambio de que $u_{ij}, v_{ij} \geq 0$.

Factorización no negativa de matrices

Es posible encontrar matrices U y V donde

- Como antes (casi) $X \sim UV$ donde U y V son matrices positivas.
- Como antes (casi) es posible reducir la dimensionalidad descartando columnas (filas) de U (V).
- Se pierde la ortogonalidad (¡necesariamente!).

Interpretación:

- Las filas de V siguen siendo preferencias.
- Las filas de U son los pesos que un cliente asigna a las distintas preferencias.
- ¡Qué bien si hubiese muchos ceros!

Algoritmos y herramientas

Algoritmos:

- NMF es un problema NP-completo (me dicen).
- Existen algoritmos heurísiticos, aproximados, etc.

Herramientas:

- Paquete NMF de R (para matrices pequeñas y medianas).
- GraphLab para matrices grandes en clústers (¡buena suerte compilando!)
- GraphChi para matrices algo menos grandes (en una única máquina).

De números positivos a probabilidades

Si U y V son positivas, pueden encontrarse matrices D, $\tilde{U}$ y $\tilde{V}$ tales que

$$UV = D\tilde{U}\tilde{V}$$

donde

- D es una matriz diagonal y $d_i \geq 0$.
- $\tilde{U}$ y $\tilde{V}$ son matrices positivas.
- Cada fila de $\tilde{U}$ suma 1.
- Cada fila de $\tilde{V}$ suma 1.

Y valores positivos que suman 1 son... ¡probabilidades!
(Nota: en lo sucesivo ignoraremos las tildes de U y V.)

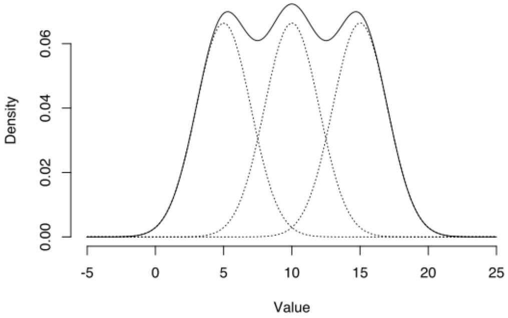
Una mezcla (mixture) de multinomiales

- La estructura probabilística revelada es la de una mezcla (*mixtura*) de multinomiales.
- Las filas de V son un menú de preferencias.
- Estas preferencias son distribuciones multinomiales con probabilidades dadas por los valores de esas filas.
- Las filas de U son los pesos de la mezcla.
- Los valores de la diagonal de D son el número (¿entero?) de muestras extraídas de cada cliente.

Mezclas de variables aleatorias

Dadas variables aleatorias $X_1, \ldots, X_n$, una mezcla de ellas con pesos $p_1, \ldots, p_n$ donde $p_i > 0$ y $\sum p_i = 1$ es otra variable aleatoria que se muestrea así:

1. Se elige una v.a. de entre las $X_1, \ldots, X_n$ con probabilidad dada por los $p_1, \ldots, p_n$.
2. Se obtiene un valor de ella.



Fuente: Wikipedia http://en.wikipedia.org/wiki/Mixture_distribution

Interpretabilidad y más

1. Tenemos una lista de preferencias comunes a los clientes
2. Son tanto o más interpretables cuantos más ceros tengan, cuanto más *puras* sean
3. Los clientes se clasifican de acuerdo con su (desigual) apego a ellas

Otras cuestiones:

1. Cuando aparece un cliente nuevo, es posible asignarle unas preferencias (i.e., sus pesos)
2. ¿Qué pasaría con clientes de los que no tienes información en absoluto? ¿Te treverías con un enfoque bayesiano?

NMF... ¿LDA para pobres?

¿Latent Dirichlet Allocation?

1. Un documento contiene palabras y trata uno o más temas.
2. Las palabras aportan información sobre el tema al que se refiere el documento.
3. Es posible construir algoritmos que detectan temas o que permiten identificar los temas a los que se refiere un documento.
4. Es una técnica bayesiana donde las distribuciones *a priori* siguen una Dirichlet.
5. ¡Es terriblemente exigente computacionalmente!

El modelo *generativo* que hemos planteado...

1. ... es conceptualmente similar al que subyace a LDA.
2. ¡Es mucho más liviano!

¡Y eso fue todo!

¿Preguntas?