Non-Negative Matrix Factorization

Overview

- What is NMF?
- Popular Applications
  - Document Clustering
- Deeper Dive into Mechanics

What is NMF?

Non-negative Matrix Factorization (NMF)

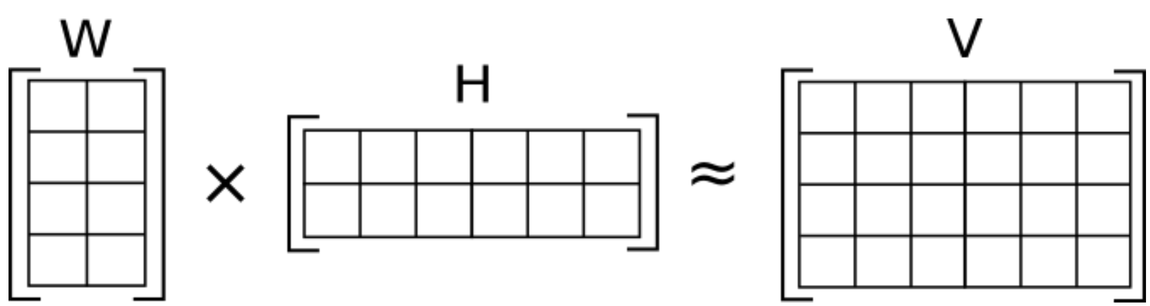
Matrix $V_{m \times n}$ where each entry $v_{ij} \geq 0$

$$\underset{m \times \textcolor{blue}{r}}{W} * \underset{\textcolor{blue}{r} \times n}{H} = \underset{m \times n}{V}$$

also $w_{ij} \geq 0$

$h_{ij} \geq 0$

- Cannot be solved analytically, so approximated numerically
- **r** set by user;
  - $r < \min(m,n)$



Matrix $V_{m \times n}$ where each entry $v_{ij} \geq 0$

$$\underset{m \times n}{V} = \underset{m \times r}{W} * \underset{r \times n}{H}$$

also $w_{ij} \geq 0$

$h_{ij} \geq 0$

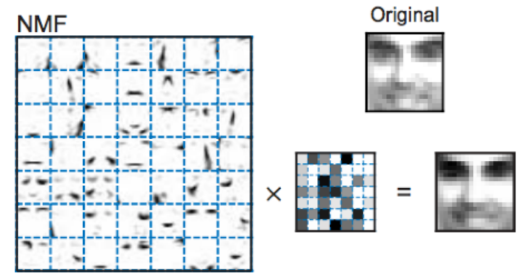
- Notice the columns of V are sum of columns of W weighted by corresponding column in h_i $$v_i = W * h_i$$

- NMF is a relatively new way of reducing dimensionality of data into linear combination of bases
  - Columns of W as basis, weighted by h_i

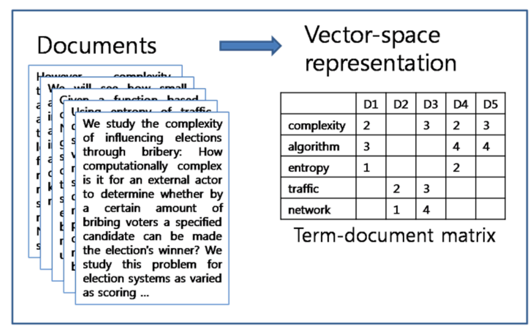
- Non-negativity constraint
  - Unlike the decompositions we've looked at thus far

Popular Applications

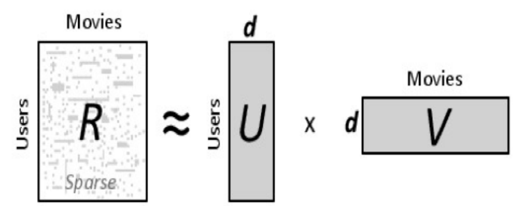
- Computer Visioning
  - Identify / classifying objects
  - Generally reducing feature space of images



- Document Clustering
  - This afternoon!



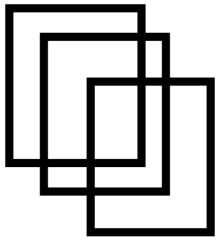
- Recommender systems
  - In just 2 days!



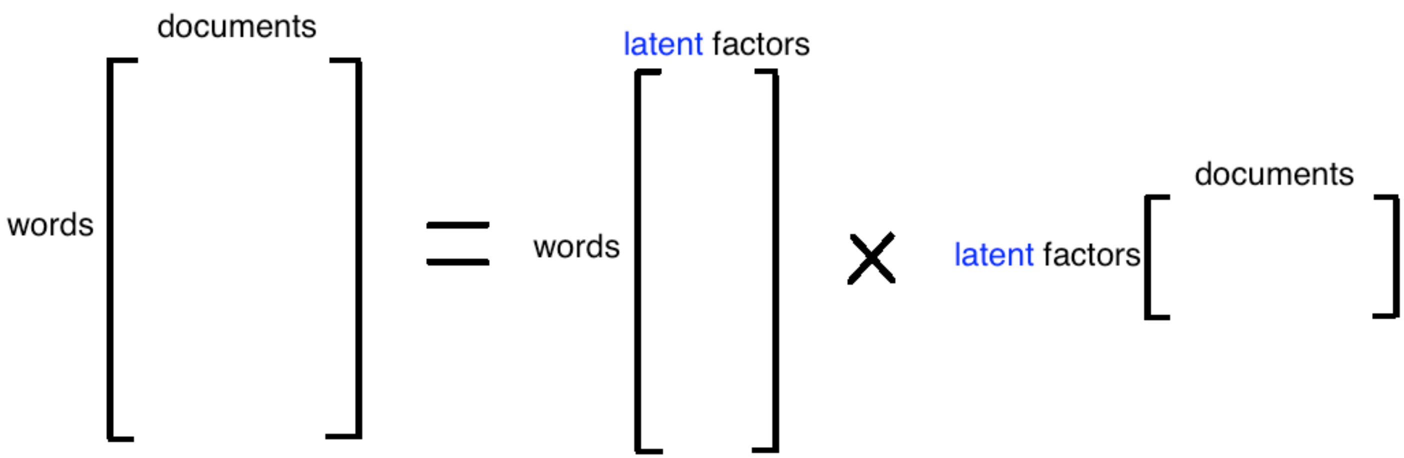
Document Clustering with NMF

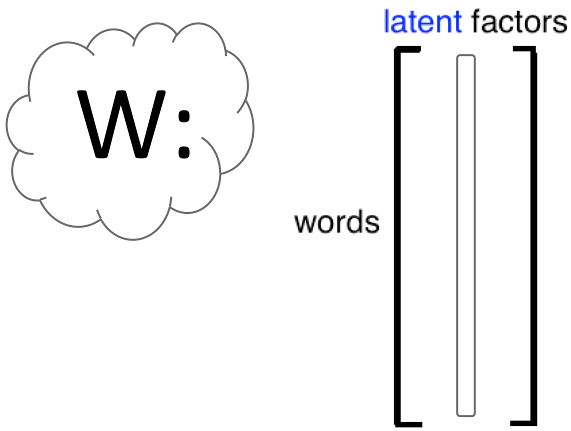
500 documents

10,000 words

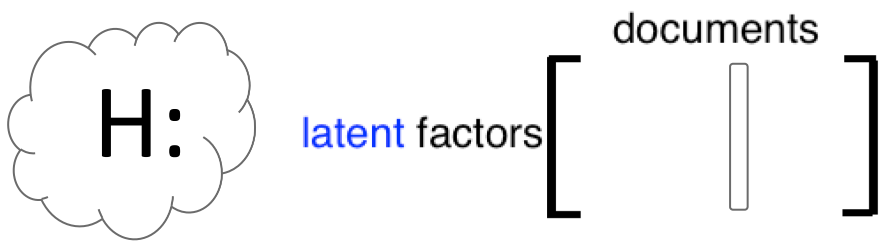


$$V \quad = \quad W \; * \; H$$

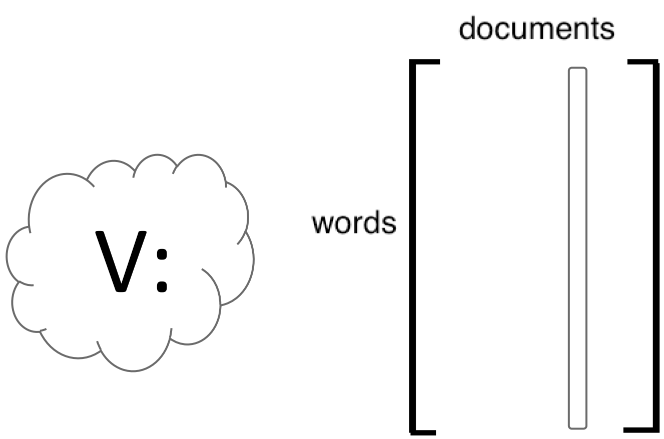




Think of column of W as document archetype where the higher the word's cell value, the higher the word's rank for that latent feature.



Think of column of H as the original document, where cell value is document's rank for a particular latent feature.



Recall $v_i = W * h_i$

Think of reconstituting a particular document as linear combination of "document archetypes" weighed by how important they are.

Mechanics

Minimize $||V - WH||^2$ with respect to *W* and *H*
subject to W, H ≥ 0

<u>Steps</u>

(1) Start with some random W and H

(2) Repeatedly adjust W and H to make RMSE smaller

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)_{a\mu}}{(W^T W H)_{a\mu}} \qquad W_{ia} \leftarrow W_{ia} \frac{(V H^T)_{ia}}{(W H H^T)_{ia}}$$

- Lee and Seung's popular "multiplicative update rules" offers compromise between speed and implementation.
- Gradient descent is simple but can be slow. Also convergence sensitive to choice of step size.

(3) Stop when some threshold is met

- Decrease in RMSE, # of iterations, etc.

Questions

- What parameter choice must you make before performing NMF?

- When doing document clustering using NMF…
  - What does a column in the W matrix represent?
  - What does a column in the H matrix represent?
  - How do we combine W and H to reconstitute a document in V (column in V)?

Appendix

Update rule:

a^{th} basis projection for i^{th} pixel

$$H_{au} \leftarrow H_{au} \sum_i W_{ia} \boxed{\frac{V_{iu}}{(WH)_{iu}}}$$

update a^{th} coefficient for the u^{th} face

sum over all pixels

ratio of actual to reconstructed pixel value for the u^{th} face

$$W_{ia} \leftarrow W_{ia} \sum_u \frac{V_{iu}}{(WH)_{iu}} H_{au}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}}$$

Normalize

Basic idea: multiply current value by a factor depending on the quality of the approximation.

If ratio > 1, then we need to increase denominator.

If ratio < 1, then we need to decrease denominator.

If ratio = 1, do nothing.

http://www.cs.cmu.edu/~02317/slides/lec_7.pdf#page=17

PCA	NMF
Unsupervised dimensionality reduction	Unsupervised dimensionality reduction
Orthogonal vectors with positive and negative coefficients	Non-negative coefficients
“Holistic”; difficult to interpret	“Parts-based”; easier to interpret
Non-iterative	Iterative (the presented algorithm)

http://www.cs.cmu.edu/~02317/slides/lec_7.pdf#page=17