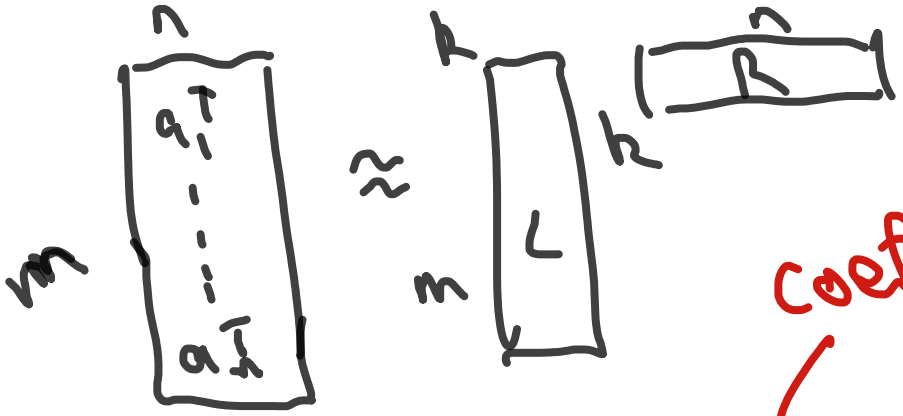
Feb 7, 2019

Linear dimensionality redux

$$A \approx L M R$$



coefficients

basis vectors

$$a_j^T \approx \sum_{r=1}^{k} L[j,k]\, r_k^T$$

$$A \approx U_k \Sigma_k V_k^T$$

basis vectors

dim. of max variance

PCs → coeffs

soln to

$$\min \|A - X\|_F \quad \text{s.t. } \operatorname{rank}(X) = k$$

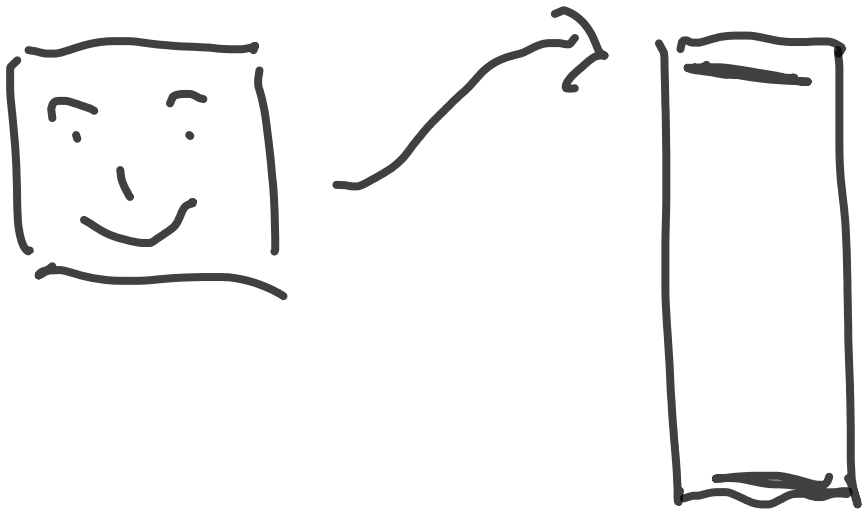
Nonnegative matrix factorization (NMF)

$$A \approx LR$$

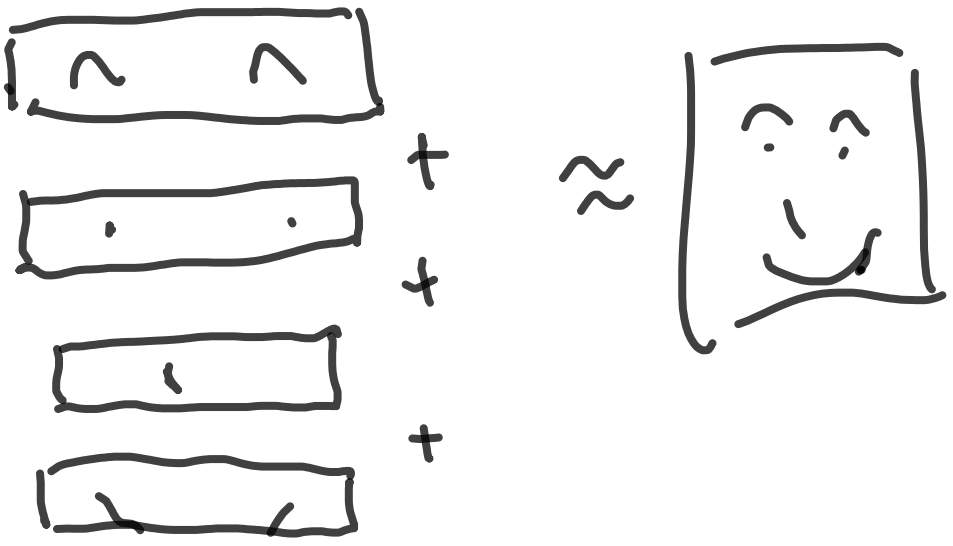
s.t. $L, R \geq 0$

$A \geq 0$ (given)

Example: face images



R matrix:



- faces are made up of a few parts
- Combinations are positive

Example: text mining

topic modelling

a_j^T = text in document j

$$a_j^T \approx \sum_{\ell}^{k} L[j,\ell]\, r_\ell^T$$

$L[j,\ell]$: relevance of topic ℓ to document j

r_ℓ^T: ℓ^{th} topic

Example: graph clustering

graph A

a_j^T = connections of node j

$$a_j^T \approx \sum_{l=1}^{k} L[j, l]\, r_l^T$$

$L[j, l]$: importance of cluster l to node j

r_l^T: representative vectors (clusters)

Some NMF problems

① NMF is NP-hard

Vavasis '09

② NMF is ill-posed

$A \approx LR$

$= LBB^{-1}R$

$B = aI$, $a > 0$

$L \geq 0$, $B^{-1}R \geq 0$

Could add

- sparsity penalty
- size penalty

$$\min_{L,R} \|A - LR\|^2 + \gamma \|L\|_F^2 + \lambda \|R\|_F^2$$

$$\text{s.t. } \operatorname{rank}(L) = \operatorname{rank}(R) \leq k$$

Optimization algs for finding L, R

1. Alternating Least Squares (ALS) + projection

(i) Fix R

$$\hat{L} = \min_{L} \|A - LR\|_F^2$$

$$= \min_{L} \|A^T - R^T L^T\|_F^2$$

$$= \min_{L} \sum_{l=1}^{k} \|A^T[:, l] - R^T L^T[:, l]\|_2^2$$

LLS for

$$\hat{L}, \quad L \leftarrow \max(\hat{L}, 0)$$

(ii) fix L

$$\hat{R} = \min_{R} \|A - LR\|_F^2$$

LLS again: $R \leftarrow \max(\hat{R}, 0)$

Problems:

- need not converge
- error need not decrease

Good things

- super simple

---

② constrained ALS

(i) Fix R

$$L^* = \min_{L} \|A - LR\|_F^2 \quad \text{s.t.} \quad L \geq 0$$

$\|A - LR\|_F^2$: cvx, smooth

$L \geq 0$: cvx (linear)

(ii) Fix L

$$R^* = \min_{R} \|A - LR\|_F^2 \quad \text{s.t.} \quad R \geq 0$$

Good property:

* error nonincreasing

(Near-) separable NMF

Idea: some rows of A are "representative"

Donoho & Stodden '03

$A \approx L\, A[K_1 :]$

$K \subseteq \{1, \ldots, m\}$

When?

Topic modeling: $\geq$ one doc. on exactly one topic for each topic

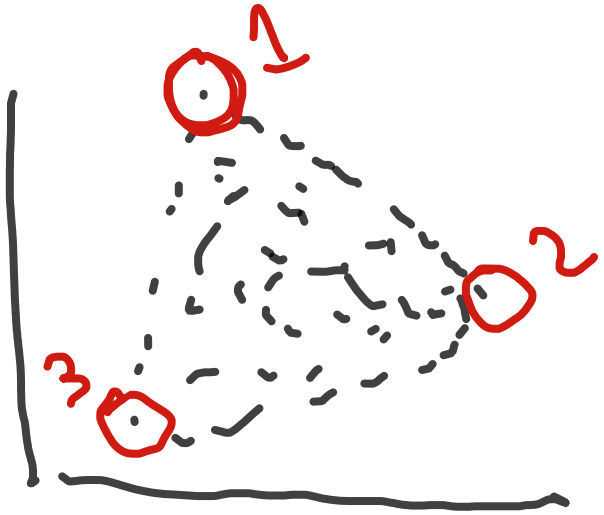
Alg: Successive Projection Algorithm

Run pivoted QR on A^T

first k picked $\Rightarrow K$

$$\min_R \|A - [A[K,:]]\|_F^2$$

$$\text{s.t. } R \geq 0$$



Separability can make NMF easier

Theorem (Arora+)

if $A = L\,A[K, :]$

$|K| = r$

Can compute K in $\text{poly}(m, n, r)$ time