Feb 13, 2020

Dimensionality reduction so far

PCA: $\|A - X\|_F^2$ s.t. X rank k

Robust PCA: $\|A - (X+S)\|_F^2$ X rank k, S sparse

Matrix completion: $\|P_\Omega(A - X)\|_F^2$ X rank k

coefficients

basis vectors

$$a_j^T \approx \sum_{s=1}^{k} L_{js}\, r_s^T$$

basis vectors of maximal variance

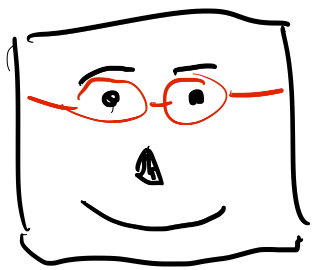
Example: PCA $\quad A \approx \underbrace{U_k \Sigma_k}_{\text{PCs}} V_k^T$

Structure: $A \geq 0$ (nonnegative)

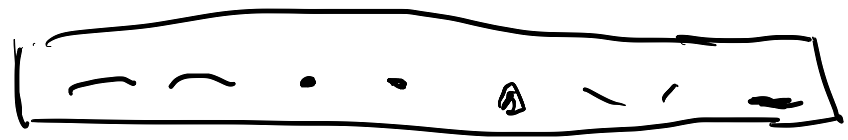
Dim redux: $A \approx LR \quad L, R \geq 0$

Nonnegative matrix factorization (NMF)

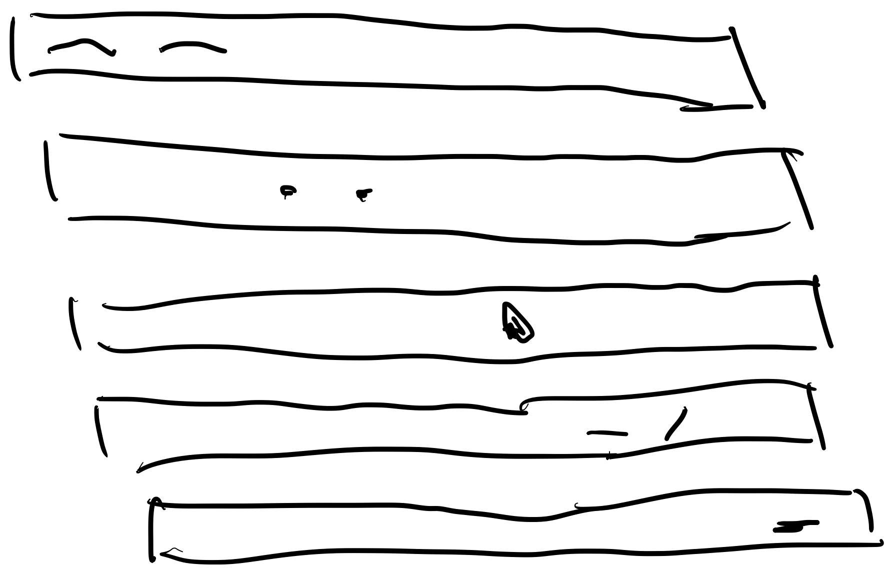
Example: face images



data point $\rightarrow$



R matrix:



basis vectors:
interpretable
nonnegative

Example: text mining

a_j^T = word distribution of doc. j

$$a_j^T \approx \sum_{s=1}^{k} L_{js} r_s^T$$

distribution of words on topic s

how much for each topic

Example: social networks

a_j^T = friends of person j

$$a_j^T \approx \sum_{s=1}^{k} L_{js} r_s^T$$

community s members

community membership

Some problems ($A \geq 0$)

approx / heuristics
restrictions

1. Given k, $\min \|A - LR\|_F^2$ s.t. $L, R \geq 0$ NP-hard (Vavasis 09)

2. ill-posed $A \approx LR = LDD^{-1}R$ $D > 0$ diagonal

add regularization

3. best rank k not nested

What to do?

$$\|A - LR\|_F^2 + \lambda(\|L\|_F^2 + \|R\|_F^2)$$

How to optimize?

Alternating leasts squares + projection (ALS)

$$\min_{L,R} \|A - LR\|_F^2 + \lambda(\|L\|_F^2 + \|R\|_F^2) \text{ s.t. } L, R \geq 0$$

(i) Fix L, opt R

$$\hat{R}^{k+1} = \min_R \|A - LR\|_F^2 + \lambda\|R\|_F^2 \quad \cancel{\text{s.t. } R \geq 0}$$

$$= \sum_{s=1}^{n} \|A(:,s) - LR(:,s)\|_2^2 + \lambda\|R(:,s)\|_2^2$$

reg LLS

$$R^{k+1} = \max(\hat{R}^{k+1}, 0)$$

Problems
- may not converge
- error need not decrease

(ii) Fix R, opt L

$$\hat{L}^{k+1} = \min_L \|A - LR\|_F^2 + \lambda\|L\|_F^2$$

$$= \min_L \|A^T - R^T L^T\|_F^2 + \lambda\|L\|_F^2$$

$$L^{k+1} = \max(\hat{L}^{k+1}, 0)$$

Constrained ALS

smooth   smooth

$$R^{k+1} = \min_R \|A - LR\|_F^2 + \lambda \|R\|_F^2$$

cvx

$$\text{s.t. } R \geq 0$$ ← linear convex

$$R_1 \geq 0 \quad R_2 \geq 0 \qquad \alpha R_1 + (1-\alpha) R_2 \geq 0$$

$$\sum_s \left\| \begin{pmatrix} A(:,s) \\ 0 \end{pmatrix} - \begin{pmatrix} L \\ \sqrt{\lambda} I \end{pmatrix} R(:,s) \right\|_2^2 \quad \text{s.t. } R(:,s) \geq 0$$

$\|Mx - b\|_2^2$ s.t. $x \geq 0$ nonnegative LS

Good property: error nonincreasing

Projected SGD

$$\min_{L,R} \frac{1}{nm}\left[\|A - LR\|_F^2 + \lambda\left(\|L\|_F^2 + \|R\|_F^2\right)\right] \quad \text{s.t. } L, R \geq 0$$

$$= \frac{1}{nm} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} f_{ij}(L, R) \longleftarrow (a_{ij} - \ell_i^T r_j)^2 + \frac{\lambda}{m}\|\ell_i\|_2^2 + \frac{\lambda}{n}\|r_j\|_2^2$$

$$\nabla_{\ell_i} f_{ij} = -2(a_{ij} - \ell_i^T r_j) r_j^T + \frac{2\lambda}{m}\ell_i$$

$$\nabla_{r_j} f_{ij} = -2(a_{ij} - \ell_i^T r_j)\ell_i^T + \frac{2\lambda}{n} r_j \qquad P(x) = \max(x, 0)$$

$$\ell_i^{k+1} = P\left(\ell_i^k - \alpha \nabla_{\ell_i} f_{ij}(L, R)\right)$$

Projected GD

$$\|A - LR\|_F = \|A^T - R^T L^T\|_F$$

$$\|X - WH\|_F$$

$$W = X(:, K)$$

$X(:, j) \approx$ distribution words for doc

$H \to$ selection non-linear combos of a few topics

W captures politics, sports, finance

at least one document on exactly one topic

(Near-) Separable NMF

$$X \approx WH \qquad W = X(:,K)$$

$$X\Pi^{T} = \underbrace{X(:,K)}_{W}\underbrace{[I \;\; H']\Pi^{T}}_{H}$$

if we knew K, then

$$\min_{H} \|X - X(:,K)H\|_F^2 \quad \text{s.t. } H \geq 0$$

Scale columns of X to sum to $1 \Rightarrow \bar{X}$

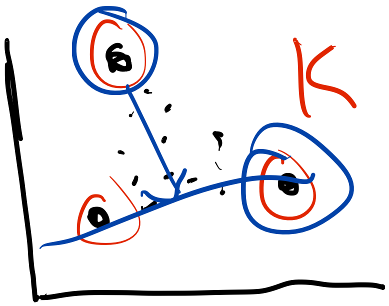
$$\bar{X} = \bar{X}(:,K)H$$

nonnegative combinations

points in a simplex

points in simplex

$$\bar{X} \subseteq \text{convex hull}(\bar{X}(:,K))$$



Successive projection alg (SPA)

Pivoted QR

for $r = 1, \dots, k$

1. Pick column t with largest 2-norm
2. add that col to K
3. $q = \bar{X}(:,t) / \|\bar{X}(:,t)\|_2$

   $\bar{X} = (I - qq^T)\bar{X}$

$$\bar{X} = \underbrace{\bar{X}(:,K)}_{W} H$$

① first col correct

$$\|X(:,j)\|_2 = \|W H(:,j)\|_2$$

strict unless $H(:,j)$ identity column

$$\leq \sum_{s=1}^{n} H(s,j) \|W(:,s)\|$$

$$\leq \max_{s} \|W(:,s)\|_2$$

② inductive step

(Gillis '14)

Near-separable $\bar{X} = \bar{X}(:,K) H + N$