Feb 5, 2019

Last time (end)

Latent Factor Models

$$A \approx L M R$$

Factors say something about data (hopefully)

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

k-means: $A \approx L R$

alternating min. steps

Today: 1. PCA
2. Robust PCA
3. CUR

---

1. PCA

features (normalized)

$$A = \begin{array}{cc} & n \\ \text{samples } m & \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \end{array}$$

Direction

$z \in \mathbb{R}^n$

$\|z\|^2 = z^T z = 1$

$a_i \in \mathbb{R}^n$

LS projection of a onto z

Claim: $a^T z$

$$\min_x \|zx - a\|^2$$

$$\cancel{z^T z}\,\hat{x} = z^T a$$

$$z = e_k \qquad e_k^T a = a_k \qquad \begin{bmatrix} 0 \\ \vdots \\ a_k \\ \vdots \\ 0 \end{bmatrix}$$

Var. of proj. onto z:

$$\max \sum_{i=1}^{m} (a_i^T z)^2$$

$$= \max \|Az\|^2$$

$$(\text{s.t. } \|z\|^2 = 1)$$

$$\|V^T z\|^2 = 1 \iff \|z\|^2 = 1$$

$$\|U \Sigma V^T z\|^2 = \|\Sigma V^T z\|$$

$$= \|\Sigma y\| \qquad y = V^T z$$

$$= \sum_{i=1} \sigma_i^2 y_i^2$$

σ_i ordered by value

$\sigma_i \geq 0$

Solution is $y = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1$

$y = V^T z$

first col of

$\Rightarrow z = Vy = Ve_1 = v_1$

Project onto data:

$Av_1 \leftarrow$ first principal component

$v_2 =$ max variance direction z_2

s.t. $z_2^T v_1 = 0$

$AV_k \Rightarrow$ projection of data

into $\mathbb{R}^k$

$$A \cdot V_k = U \Sigma V^T V_k$$

$$= U \Sigma \begin{bmatrix} I \\ 0 \end{bmatrix}$$

$$= U_k \Sigma_k$$

---

② Robust PCA

$$A \approx LMR$$

$$\underbrace{U_k \Sigma_k}_{\text{PCs}} V_k^T$$

Problems: measurement error
outliers
data entry
(not too many)

idea: $A \approx \underbrace{LR}_{X} + \underbrace{S}_{\text{sparse}}$

① $\underset{X,S}{\text{minimize}} \quad \operatorname{rank}(X) + \gamma \operatorname{nnz}(S)$

s.t. $\quad X + S = A$

② $\underset{X,S}{\min} \quad \operatorname{rank}(X) + \gamma \operatorname{nnz}(S) + \lambda \|A - X - S\|_F$

$\mathrm{rank}(X) = \mathrm{nnz}([\sigma_1, \ldots, \sigma_n])$

cvx surrogate

$$\left\| \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{bmatrix} \right\|_1$$

$$= \|X\|_*$$ nuclear norm

$\mathrm{nnz}(S)$ cvx surrogate

$$\|\mathrm{vec}(S)\|_1$$

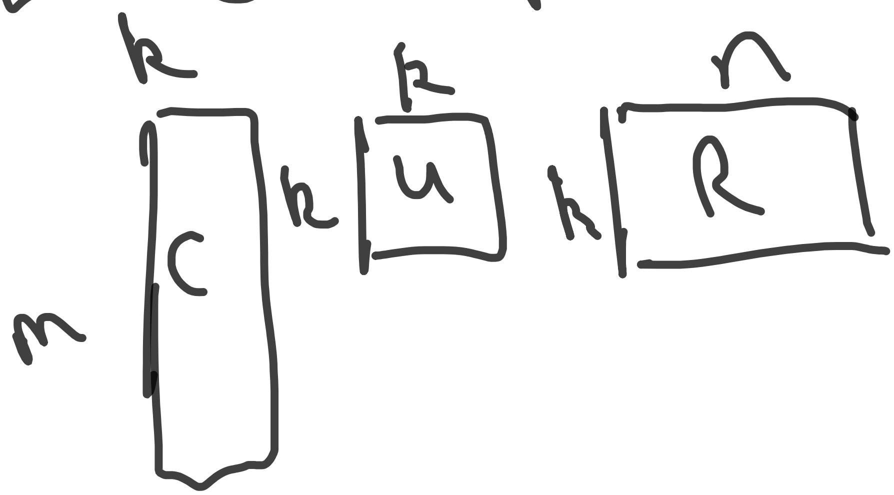
$$\boxed{\begin{aligned} \min \quad & \|X\|_* + \gamma \|\mathrm{vec}(S)\|_1 \\ \text{s.t.} \quad & L + S = A \end{aligned}}$$

(ADMM can solve efficiently)

3 CUR

problem: low-rank part may not be interpretable

$A \approx C\,U\,R$



C is k cols of A

R is k rows of A

0 Trun. SVD $A \approx U_k \Sigma_k V_k^T$ $O(mk^2)$

Alg (Mahoney + Drineas '09)

ColSelect($A^{\boxed{m \times n}}$, k, ε, c)

(i) $\pi_j = \frac{1}{k} \sum_{i=1}^{k} [v_i(j)]^2$ $O(nk)$ $O(mk)$

S $\begin{cases} \text{(ii) keep } j^{th} \text{ col of } A \text{ w.p.} & O(n) \\ p_j = \min(1, c\pi_j) & O(n) \\ [c = O(k \log k / \varepsilon^2)] & \end{cases}$

(iii) return S

CUR: ① S = ColSelect(A, k, ε, c) $O(mk^2)$

② T = ColSelect(A^T, k, ε, c) $O(mk^2)$

3 $C = A[:, S]$

$R = A[T, :]$

$m \geq n$

$O(mk^2)$

$U = C^+ A R^+$

$O(|S|^2 \cdot n)$ $O(|T|^2 \cdot n)$

$A \approx CUR$

$M = U \Sigma V^T$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

$$M^+ = V \begin{pmatrix} 1/\sigma_1 & & & & & \\ & \ddots & & & & \\ & & 1/\sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} U^T$$

$A = CUR$

Theorem:

$$\|A - CUR\|_F \le (2+\varepsilon) \underbrace{\|A - U_k \Sigma_k V_k^T\|_F}_{\text{OPT}}$$

w.h.p.