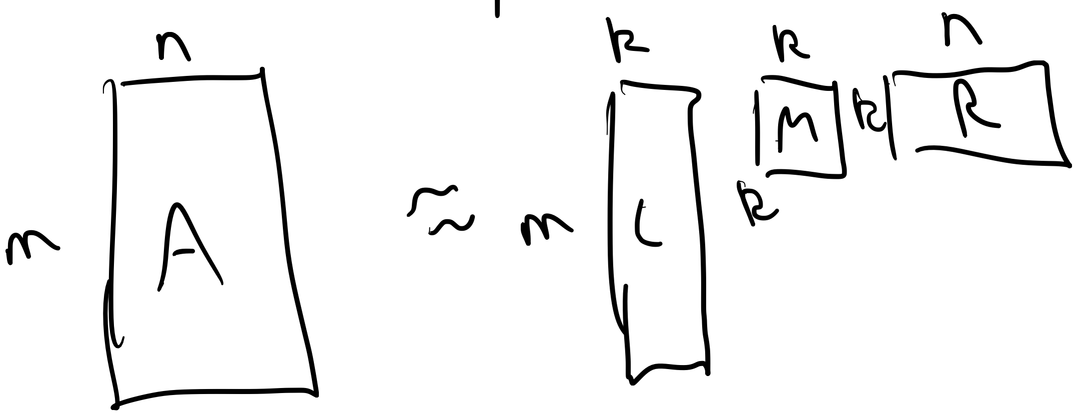
Feb 6, 2020

Kargbo OH: Rhodes 405
Mon 2:30

Dimensionality redux / latent factor models



Goal:

- generative models
- think in smaller dims
- factors say something meaningful
- denoising

k-means: $A = LR$

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

$$\min_Z \|A - Z\|_F$$

s.t. $\text{rank}(Z) = k$

Principal component analysis (PCA)

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

Direction $z \in \mathbb{R}^n$ that "summarizes" the data the most

$\|z\|_2^2 = 1$

$a_i \in \mathbb{R}^n$ LS proj onto z?

$$\hat{x} = \underset{x \in \mathbb{R}^1}{\operatorname{argmin}} \|zx - a_i\|^2 \qquad \cancel{z^T z}\,\hat{x} = z^T a_i$$

$(z = e_k \Rightarrow e_k^T a_i = (a_i)_k)$

(assume)

mean of projection: $\frac{1}{n}\sum_{i=1}^{n}(a_i^T z)z = \frac{1}{n}\underbrace{\left(\sum_{i=1}^{n} a_i\right)^T}_{=0} z\, z$

var of projection: $\frac{1}{n}\sum_{i=1}^{n}\left[(a_i^T z)z\right]^T\left[(a_i^T z)z\right]$

$$= \frac{1}{n}\sum_{i=1}^{n}(a_i^T z)^2 \cancel{z^T z} = \frac{1}{n}\|Az\|_2^2$$

maximize var $\Rightarrow$ $\max_{z^T z = 1} \|Az\|_2^2$

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

$\|y\|_2^2 = 1 \Leftrightarrow \|z\|_2^2 = 1$

$\max_y \|\Sigma y\|^2$

$= \sum_{i=1}^{n} \sigma_i^2 y_i^2 \quad \sigma_1 \geq \dots \sigma_n \geq 0$

$y = e_1 \Rightarrow z = Ve_1 = v_1$

Av_1 is first principal component

$= U\Sigma V^T v_1 = U\Sigma e_1 = \sigma_1 u_1$

Generalize: find z_2 orthog to v_1 that maximizes variance

$$\max_{z_2} \|Az_2\|^2 \quad \text{s.t.} \quad z_2^T v_1 = 0 \quad z_2^T z_2 = 1$$

$$z_2 = v_2$$

$$\max_{Z \in \mathbb{R}^{n \times k}} \|AZ\|_F \quad \text{s.t.} \quad Z^T Z = I$$

$$Z = V_k \quad \text{projection} \quad AV_k = U_k \Sigma_k$$

first k PCs

"variance explained"

$$\frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_n^2}$$

Possible problems: data corruption, measurement error, outliers

idea: $\boxed{A \approx LR + S}$ Robust PCA

sparse to adjust for

$$\sum_{ij} |S_{ij}|$$

$$Z = LR$$

First model: $$\min \frac{1}{2} \| A - (Z+S) \|_F^2 + \lambda \|S\|_1$$
$$\text{s.t. } \operatorname{rank}(Z) = k \quad \cancel{\operatorname{nnz}(S) \le r}$$

How to optimize?

1. Fix S, solve for Z
2. Fix Z, solve for S

$$\min \frac{1}{2}\|A-(Z+S)\|_F^2 + \lambda\|S\|_1 \quad \text{s.t. } \operatorname{rank}(Z) = k$$

① Fix S, solve for Z

$$Z^{k+1} = \min_Z \|(A-S^k)-Z\|_F^2 \quad \text{s.t. } \operatorname{rank}(Z)=k$$

$$Z = U_k \Sigma_k V_k^T \text{ for } A-S$$

② Fix Z, solve for S

$$\min_S \frac{1}{2}\|(A-Z)-S\|_F^2 + \lambda\|S\|_1$$

$$\operatorname{prox}_f(v) = \arg\min_x \left(f(x) + \frac{1}{2}\|x-v\|_2^2 \right)$$

$$f = \lambda\|\cdot\|_1 \quad x = \operatorname{vec}(S) \quad v = \operatorname{vec}(A-Z)$$

$$\mathrm{prox}_{\lambda\|\cdot\|_1}(v) = \underset{x}{\mathrm{argmin}} \left(\lambda \|x\|_1 + \frac{1}{2}\|x-v\|_2^2 \right)$$

$$0 \in \nabla_x \frac{1}{2}\|x-v\|_2^2 + \partial \lambda \|x\|_1$$

$$\left(\partial f(x) = \{ g \mid f(z) \geq f(x) + g^T(z-x) \ \forall z \} \right)$$

$$x_i^* > 0 : \quad 0 = x_i^* - v_i + \lambda \Rightarrow x_i^* = v_i - \lambda \quad (v_i > \lambda)$$

$$x_i^* < 0 : \quad 0 = x_i^* - v_i - \lambda \Rightarrow x_i^* = v_i + \lambda \quad (v_i < \lambda)$$

$$x_i^* = 0 \qquad 0 = -v_i + \lambda[-1, 1]$$

$$x_i^* = 0 \iff v_i \in [-\lambda, \lambda]$$

$$x_i^* = \begin{cases} v_i - \lambda\,\mathrm{sign}(v_i) & |v_i| > \lambda \\ 0 & |v_i| \leq \lambda \end{cases}$$

$$= S_\lambda(v)$$

$$\min_{S} \frac{1}{2}\|(A-Z^{k}) - S\|_F^2 + \lambda\|S\|_1$$

$S^{k+1} =$

$$S^{k+1} = S_\lambda(A-Z)$$

$$Z^{k+1} = TSVD(A-S)$$

$$S_{ij}^{k+1} = \begin{cases} B_{ij} - \lambda\,\mathrm{sign}(B_{ij}) & |B_{ij}| > \lambda \\ 0 & |B_{ij}| < \lambda \end{cases}$$

$$B = A - Z$$

$$\min_{Z,S} \quad \overset{\|Z\|_*}{\cancel{\operatorname{rank}(Z)}} + \gamma\, \cancel{\operatorname{nnz}(S)}\ \|S\|_1$$

$$\text{s.t.} \quad Z + S = A$$

$$\|A\|_* = \operatorname{trace}\left(\sqrt{A^T A}\right) = \sum_{i=1}^{n} \sigma_i$$

$$\|A\|_* = \left\| \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{pmatrix} \right\|_1$$

$$\min_{Z,S} \quad \|Z\|_* + \gamma \|S\|_1 + \frac{1}{2\beta}\|A - Z - S\|_F^2$$

$$\text{s.t.} \quad Z + S = A$$

$$L(Z, S, \Lambda) = \|Z\|_* + \gamma\|S\|_1 + \frac{1}{2\beta}\|A - (Z+S)\|_F^2$$
$$+ \langle \Lambda, Z + S - A \rangle$$

Alternate $S / Z / \Lambda$

$$S^{k+1} = \arg\min_S \; \beta\gamma\|S\|_1 + \frac{1}{2}\|(A - Z^k) - S\|_F^2$$
$$+ \langle \beta\Lambda^k, Z^k + S - A \rangle$$

$$= \arg\min_S \; \beta\gamma\|S\|_1 + \frac{1}{2}\|A - Z^k - \beta\Lambda^k - S\|_F^2$$

$=$

Alternate $S / Z / \Lambda$

$$S^{k+1} = \arg\min_{S} \beta\gamma \|S\|_1 + \frac{1}{2}\|(A - Z^k) - S\|_F^2 + \langle \beta \Lambda^k, Z^k + S - A \rangle$$

$$= \arg\min_{S} \beta\gamma \|S\|_1 + \frac{1}{2}\|A - Z^k - \beta\Lambda^k - S\|_F^2$$

$$= S_{\beta\gamma}\left(A - Z^k - \beta\Lambda^k\right)$$

$$Z^{k+1} = \arg\min_{Z} \ \beta\|Z\|_* + \frac{1}{2}\|(A - S^{k+1}) - Z\|_F^2 + \langle \beta\Lambda^k, Z + S^{k+1} - A\rangle$$

$$= \arg\min_{Z} \ \beta\|Z\|_* + \frac{1}{2}\|A - S^{k+1} + \beta\Lambda^k - Z\|_F^2$$

$$\mathrm{prox}_{\beta\|\cdot\|_*}\left(\underbrace{A - S^{k+1} + \beta\Lambda^k}_{X}\right)$$

$$X = U\Sigma V^T \qquad s_\beta(\sigma) = \max(\sigma - \beta, 0)$$

$$U s_\beta(\Sigma) V^T = U_k \Sigma_k V_k^T$$

$$\langle \Lambda, Z^{k+1} + S^{k+1} - A \rangle$$

$$\Lambda^{k+1} = \Lambda^{k} - \frac{1}{\beta}\left(Z^{k+1} + S^{k+1} - A\right)$$

ADMM