9/13/23

Lecture 3: Iterative methods for tensors

Whitening:

Given: $M = \sum \lambda_i u_i u_i^T$ for u_i's non-orthogonal

$T = \sum \lambda_i u_i^{\otimes 3}$

Let $\underset{d\times k}{V}\,\underset{k\times k}{D}\,\underset{k\times d}{V^T}$ be eigendecomposition of M, define

$$W \triangleq V D^{-1/2} \in \mathbb{R}^{d\times k}$$

$$\hat{u}_i \triangleq \lambda_i W^T u_i \in \mathbb{R}^k$$

Motivation: W transforms M into Id_k!

$$W^T M W = D^{-1/2} \underbrace{V^T V}_{Id_k} D V^T V D^{-1/2}$$

$$= D^{-1/2} D D^{-1/2} = Id_k$$

$$\shortparallel$$

$$= \sum \lambda_i (W^T u_i)(W^T u_i) = \sum \hat{u}_i \hat{u}_i^T$$

So $\hat{u}_i$'s are orthonormal basis for $\mathbb{R}^k$.

Then consider

$$T' \triangleq T(W, W, W) \in \mathbb{R}^{k \times k \times k}$$

where $T'(x,y,z) \triangleq T(Wx, Wy, Wz)$

Then $T'(x,y,z) = \sum_i \lambda_i \langle Wx, u_i \rangle \langle Wy, u_i \rangle \langle Wz, u_i \rangle$

$$= \sum_i \lambda_i (x^T W^T u_i) \cdot (y^T W^T u_i) \cdot (z^T W^T u_i)$$

$$= \sum_i \lambda_i \left(\lambda_i^{-1/2} \langle \hat{u}_i, x \rangle \right) \cdot \left(\lambda_i^{-1/2} \langle \hat{u}_i, y \rangle \right) \cdot \left(\lambda^{-1/2} \langle \hat{u}_i, z \rangle \right)$$

$$= \sum_i \lambda_i^{-1/2} \langle \hat{u}_i, x \rangle \langle \hat{u}_i, y \rangle \langle \hat{u}_i, z \rangle$$

So $\boxed{T' = \sum_i \lambda_i^{-1/2} \hat{u}_i^{\otimes 3}}$,

i.e. T' has orthogonal components.

Tensor power method for non-orthogonal components:

(without whitening) [Vatsal-Sharan '17]

$$T = \sum_{i=1}^{k} u_i^{\otimes 3}$$

Iterates of tensor power method (z_t) given by:

$$z_t' \triangleq T(z_{t-1}, z_{t-1}, :)$$

$$= \sum_{i=1}^{k} \langle z_{t-1}, u_i \rangle^2 u_i$$

$$z_t \triangleq z_t' / \|z_t'\|$$

Define $a_{i,t} \triangleq \langle u_i, z_t \rangle$, $\hat{a}_{i,t} \triangleq \dfrac{a_{i,t}}{a_{1,t}}$

$$a_{j,t} = \frac{\sum_i a_{i,t-1}^2 \overbrace{\langle u_i, u_j \rangle}^{c_{i,j}}}{\|z_t'\|} = \frac{a_{1,t-1}^2 \sum_i \hat{a}_{i,t-1}^2 c_{i,j}}{\|z_t'\|}$$

Because $\|z_t'\|$ is fixed factor independent of j,

$$\boxed{\hat{a}_{j,t} = \frac{\sum_i \hat{a}_{i,t-1}^2 c_{i,j}}{\sum_i \hat{a}_{i,t-1}^2 c_{i,1}}}$$

Rewriting this further to isolate $i=1$ terms,

$$\hat{a}_{j,t} = \frac{c_{1,j} + \sum_{i \neq 1} \hat{a}^2_{i,t-1} c_{i,j}}{1 + \underbrace{\sum_{i \neq 1} \hat{a}^2_{i,t-1} c_{i,1}}_{|\cdot| \leq kc_{max} << 1}}$$

$$\simeq \left(c_{1,j} + \sum_{i \neq 1} \hat{a}^2_{i,t-1} c_{i,j}\right)\left(1 - \sum_{i \neq 1} \hat{a}^2_{i,t-1} c_{i,1}\right) \quad (1)$$

We show

$$\max_{j \neq 1} |\hat{\alpha}_{j,t}| < \beta_t \quad (\delta)$$

for sequence (β_t) defined recursively by

$$\boxed{\begin{aligned} \beta_0 &= \max_{j \neq 1} |\hat{\alpha}_{j,0}| \\ \beta_t &= c_{max} + \beta^2_{t-1} + 3kc_{max}\beta^2_{t-1} \end{aligned}}$$

Provided $kc_{max} << 1 - \beta_0$, can show $\beta_t \leq 1$
(proof omitted)

<u>Pf of (δ)</u> :

Note

$$\left|c_{1,j} + \sum_{i \neq 1} \hat{a}^2_{i,t-1} c_{i,j}\right| = \left|c_{1,j} + \hat{a}^2_{j,t-1} + \sum_{i \neq j,1} \hat{a}^2_{i,t-1} c_{i,j}\right|$$

$$\leq c_{max} + \beta^2_{t-1} + kc_{max}\beta^2_{t-1}$$

So by (1),

$$|\hat{a}_{j,t}| < (c_{max} + \beta_{t-1}^2 + k c_{max} \beta_{t-1}^2)(1 + k c_{max} \beta_{t-1}^2)$$
$$\leq c_{max} + \beta_{t-1}^2 + 2kc\beta_{t-1}^2 < \beta_t . \quad \square$$

Then suffices to analyze the recursion defining β_t (note: we have reduced keeping track of $k-1$ different quantities $\hat{a}_{2,t}, \ldots, \hat{a}_{k,t}$ to just keeping track of a single quantity!).

To get intuition, consider case where A_i's orthogonal, i.e. $c_{max} = 0$.

Then $\beta_t = \beta_{t-1}^2$, so

$$\beta_t = \beta_0^{2^t}$$

i.e. if $\beta_0 < 1$, then $\beta_t \to 0$ at <u>doubly exponential</u> rate.

For $c_{max} > 0$ case, β_0 has to be sufficiently smaller than 1, i.e.

$$1 - \beta_0 \gg k c_{max}$$

If $c_{max} \ll \frac{1}{k^2}$, then this is satisfied w.h.p by randomly initializing (proof omitted, see Lemma 1 in [Sharan-Valiant], link on course page).

Analyzing recursion for (β_t):

3 stages:
1). $\beta_t \geq 0.1$
2). $0.1 \geq \beta_t \geq \sqrt{\eta}$ $\quad \eta \triangleq \max(c_{max}, 1/d)$
3) $\beta_t \leq \sqrt{\eta} \leftarrow$ ☺

<u>Stage 1</u>: note $c_{max} \leq k c_{max} \beta_{t-1}^2$
b/c $k\beta_{t-1}^2 \geq 0.1k \geq 1$
(for k larger than some constant)

so $\beta_t \le (1+4kc_{max})\beta_{t-1}^2$,

and unrolling, we get

$$\beta_t \le (1+4kc_{max})^{1+\cdots+2^{t-1}} \beta_0^{2^t}$$

$$= (1+4kc_{max})^{2^t-1} \beta_0^{2^t}$$

$$\le \left[\beta_0(1+4kc_{max})\right]^{2^t}$$

so if $\beta_0 \le 1-5kc_{max}$ (which happens w.h.p.),

$$\le (1-\underbrace{kc_{max}}_{\le \frac{1}{k^2}})^{2^t}$$

so we stay in this stage for $\lg k$ iterations

<u>Stage 2</u>: Reindex so β_0 is start of this stage

because $\beta_t \ge \sqrt{\eta} \ge \sqrt{c_{max}}$,

$$\beta_t = (1+3kc_{max})\beta_{t-1}^2 + c_{max} \le (2+3\underbrace{kc_{max}}_{<<1})\beta_{t-1}^2$$

$$\le 3\beta_{t-1}^2$$

Unrolling, we get

$$\beta_t \leq 3^{2^t - 1} \cdot \beta_0^{2^t}$$

$$\leq (3\beta_0)^{2^t}$$

$$\leq (0.3)^{2^t}$$

So we stay in this stage for $O(\lg \lg d)$ iterations. $\square$