Lect 3 end remarks about ALS.

Recap ALS algo:

Three Matricizations. $T_{(1)}$; $T_{(2)}$; $T_{(3)}$

- $T^{\alpha\beta\gamma}$ order 3. Suppose you know that rank is R.

- Initialize algo $B^{(0)}$, $C^{(0)}$ that are arrays of dim $I_2 \times R$, $I_3 \times R$.

  full column rank.

- Iterations $m \geq 0$.

$$A^{m+1} = T_{(1)} \left(C^{(m)} \otimes_{khr} B^{(m)} \right) \left(C^{mT} C^{m} * B^{mT} B^{m} \right)^{-1}$$

$$B^{m+1} = T_{(2)} \left(A^{(m+1)} \otimes_{khr} C^{(m)} \right) \left(\boxed{A^{m+1\,T} A^{m+1} * C^{mT} C^{m}} \right)^{-1}$$

$$C^{m+1} = T_{(3)} \left(B^{m+1} \otimes_{khr} A^{m+1} \right) \left(B^{m+1\,T} B^{m+1} * A^{m+1\,T} A^{m+1} \right)^{-1}$$

- This algo comes with various caveats & problems (although it is still very useful & popular).

$\longrightarrow$ Remarks

Remarks:

1) Attempt is to minimize $\| T - \sum_{r=1}^{R} \underbrace{\vec{a}_r \otimes \vec{b}_r \otimes \vec{c}_r}_{\text{unknowns}} \|_F^2$

try min over a's, b', c's.

Highly non convex and if the algo converges.

then it might end up to some local min.

Might be a good one or not.
depend on initialization.

2) The rank R of T has to be known.

3) Because at each iteration (except first step)

there is no guarantee that usual inverses exist

→ like MP pseudo-inverse exist always but not

necessarily computable from

If A^{m+1}, C^m are full col rank.

$$\left(\underbrace{(A^{m+1} \otimes_{khr} C^m)\,(A^{m+1\,T} A^{m+1} * C^{m\,T} C^m)^{-1}}_{((A^{m+1} \otimes_{khr} C^m)^T)^{+}} \right.$$

← exists.

people regularize the formulas:

$$(A^{n+1} \otimes_{khr} C^{n}) \left(A^{n+1\,T} A^{n+1} * C^{n\,T} C^{n} + \underline{\underline{\lambda I}} \right)^{-1}$$

here $\lambda > 0$ small reg parameter

and $(\text{Identity})_{R \times R} = I$.

$$\Downarrow$$

- Regularized algorithm makes sense certainly for $\lambda > 0$ through all iterations.

Implemented with $\underline{\underline{\text{small reg } \lambda > 0}}$.

4) Then what is the quantity that we attempt to minimize in principle?

$$\| T - \sum_{r=1}^{R} \vec{a}_r \otimes \vec{b}_r \otimes \vec{c}_r \|_F^2 + \lambda \sum_{r=1}^{R} \underline{\|\vec{a}_r\|_2^2 + \|\vec{b}_r\|^2 + \|\vec{c}_r\|^2}$$

Ridge regularization.

$\mathrm{Tr}\, A^T A = \|A\|_F^2$.

- ALS with regularization:

$$\underset{A}{\mathrm{argmin}} \left\{ \| T_{(1)} - A \underbrace{(C \otimes_{khr} B)^T}_{\text{"known"}} \|_F^2 + \underbrace{\lambda \|A\|_F^2}_{\text{regularization}} \right\}.$$

$$\underset{B}{\mathrm{argmin}} \left\{ \| T_{(2)} - B (A \otimes_{khr} C)^T \|_F^2 + \lambda \|B\|_F^2 \right\}.$$

$$\underset{C}{\mathrm{argmin}}\ \| T_{(3)} - C (B \otimes_{khr} A)^T \|_F^2 + \lambda \|C\|_F^2$$

→ Regularized version of ALS.

Rank R has still to be known.

still not guaranteed that you get true min

of $\|T - \sum_r \vec{a}_r \otimes \vec{b}_r \otimes \vec{c}_r\|_F^2 + \lambda \sum_r \|a_r\|^2 + \|b_r\|^2 + \|c_r\|^2$.

still not convex

→ But better behaved numerically.

// end of ALS.
Next time move to New Methods
New Algorithms.