


Lect 3 and remarks about ALS.

Recap ALS algo:

Three Matrices:

$T_{(1)}; T_{(2)}; T_{(3)}$

$T \propto \beta \gamma$

order 3.

Suppose you know
that rank is R .

Initialize algo $B^{(0)}, C^{(0)}$ that are
arrays of dim $\mathbb{I}_2 \times R, \mathbb{I}_3 \times R$.

full column rank.

Iterative $m \geq 0$.

$$\begin{aligned} A^{m+1} &= T_{(1)} \left(C^{(m)} \otimes_{\text{Kh}} B^{(m)} \right) \left(C^{mT} C^m * B^{mT} B^m \right)^{-1} \\ B^{m+1} &= T_{(2)} \left(A^{(m+1)} \otimes_{\text{Kh}} C^{(m)} \right) \left(A^{m+1T} A^{m+1} * C^{mT} C^m \right)^{-1} \\ C^{m+1} &= T_{(3)} \left(B^{m+1} \otimes_{\text{Kh}} A^{m+1} \right) \left(B^{m+1T} B^{m+1} * A^{m+1T} A^{m+1} \right)^{-1} \end{aligned}$$

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

→
Remarks

Remark 1:

- 1) Attempt is to minimize $\|T - \underbrace{\sum_{r=1}^R \vec{a}_r \otimes \vec{b}_r \otimes \vec{c}_r}_{\text{unknowns}}\|_F^2$
by min over a 's, b 's, 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 a and b -vectors exist

→ like MP pseudo-inverse exist always but not

necessarily computable from

if A^{n+1}, C^n
are full col
rank.

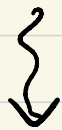
$$\left\| \underbrace{(A^{n+1} \otimes_{\text{thr}} C^n) (A^{n+1T} A^{n+1} * C^n C^n)^{-1}}_{\left((A^{n+1} \otimes_{\text{thr}} C^n)^T \right)^+} \right\| \leftarrow \text{exists.}$$

people preferize the formulae:

$$(A^{m+1} \otimes_{khr} C^m) \left(A^{m+1T} A^{m+1} \otimes C^m C^m + \lambda I \right)^{-1}$$

here $\lambda > 0$ small reg parameter

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



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

Implemented with small reg $\lambda > 0$.

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

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

Ridge regularization.

• ALS with regularization:

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

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

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

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

→ Regularized version of ALS.

Rank R has still to be known.

still not guaranteed that you get true min

$$\text{of } \|T - \sum_r \vec{c}_r \otimes \vec{h}_r \otimes \vec{c}_r\|_F^2 = \lambda \left(\sum_r \|\vec{c}_r\|^2 + \|\vec{h}_r\|^2 + \|\vec{c}_r\|^2 \right)$$

still not convex

→ But better behaved Numerically.

end of ALS.
Next time have to New Methods
New Algor.

