Computer Science Department, University of Bonn

Knowledge Graph Analysis

Solutions of Exercise Sheet 5

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann

December 20, 2018

1 In Class

1a) **Alternating Least Squares (ALS)**

We want to show that[1] $\hat{\boldsymbol{\mathcal{T}}} = \boldsymbol{\mathcal{R}} \times_1 \boldsymbol{A} \times_2 \boldsymbol{A}$ is equivalent to

$$\hat{\boldsymbol{\mathcal{T}}}_{\mathbf{(1)}} = \boldsymbol{A}\boldsymbol{\mathcal{R}}_{\mathbf{(1)}}(\boldsymbol{I} \otimes \boldsymbol{A})^T \ .$$

Using the definition of the *Kronecker Product* one gets

$$\boldsymbol{I} \otimes \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A} & 0 & \dots & 0 \\ 0 & \boldsymbol{A} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{A} \end{bmatrix}$$

and thus

$$(\boldsymbol{I} \otimes \boldsymbol{A})^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}} & 0 & \dots & 0 \\ 0 & \boldsymbol{A}^{\mathrm{T}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{A}^{\mathrm{T}} \end{bmatrix} \ .$$

[1] Note, that it is also equivalent to $\hat{\boldsymbol{\mathcal{T}}}_{\mathbf{(2)}} = \boldsymbol{A}\boldsymbol{\mathcal{R}}_{\mathbf{(2)}}(\boldsymbol{I} \otimes \boldsymbol{A})^T$ which can be shown analogously.

Recalling that $\boldsymbol{\mathcal{R}}_{\mathbf{(1)}} = [\boldsymbol{\mathcal{R}}_{\mathbf{::1}}, \boldsymbol{\mathcal{R}}_{::2}, \ldots, \boldsymbol{\mathcal{R}}_{::N_r}]$, where N_r is the number of relations or tensor frontal slices, one gets as on slide 38 of lecture 5

$$\begin{aligned}
\hat{\boldsymbol{\mathcal{T}}}_{\mathbf{(1)}} &= \boldsymbol{A}\boldsymbol{\mathcal{R}}_{\mathbf{(1)}}(\boldsymbol{I} \otimes \boldsymbol{A})^T \\
&= \boldsymbol{A}(\boldsymbol{\mathcal{R}}_{\mathbf{::1}}, \boldsymbol{\mathcal{R}}_{::2}, \ldots, \boldsymbol{\mathcal{R}}_{::N_r})(\boldsymbol{I} \otimes \boldsymbol{A})^T \\
&= [\boldsymbol{A}\boldsymbol{\mathcal{R}}_{::1}, \ldots, \boldsymbol{A}\boldsymbol{\mathcal{R}}_{::N_r}] \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}} & 0 & \ldots & 0 \\ 0 & \boldsymbol{A}^{\mathrm{T}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{A}^{\mathrm{T}} \end{bmatrix} \\
&= [\boldsymbol{A}\boldsymbol{\mathcal{R}}_{::1}\boldsymbol{A}^{\mathrm{T}}, \ldots, \boldsymbol{A}\boldsymbol{\mathcal{R}}_{::N_r}\boldsymbol{A}^{\mathrm{T}}] \ .
\end{aligned}$$

Therefore we get $\hat{\boldsymbol{\mathcal{T}}}_{:,:,k} = \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^{\mathrm{T}}$, from which follows that the equivalence above holds, since from the definition of RESCAL in the 5th lecture we already know that $\hat{\boldsymbol{\mathcal{T}}}_{:,:,k} = \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^{\mathrm{T}}$, for $k = 1, \ldots, N_r$ is equivalent to $\hat{\boldsymbol{\mathcal{T}}} = \boldsymbol{\mathcal{R}} \times_1 \boldsymbol{A} \times_2 \boldsymbol{A}$.

1b) To show that the equivalence

$$||\boldsymbol{\mathcal{T}}_{:,:,k} - \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^{\mathrm{T}}||^2 = ||\mathrm{vec}(\boldsymbol{\mathcal{T}}_{:,:,k}) - (\boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{\mathcal{R}}_{:,:,k})||^2$$

holds for

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \boldsymbol{\mathcal{R}}_{:,:,k} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \boldsymbol{\mathcal{T}}_{:,:,k} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}$$

first we calculate $\boldsymbol{\mathcal{T}}_{:,:,k} - \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^T$. Note, that

$$\boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \quad .$$

$$= \begin{bmatrix} a_{11}r_{11} + a_{12}r_{21} & a_{11}r_{12} + a_{12}r_{22} \\ a_{21}r_{11} + a_{22}r_{21} & a_{21}r_{12} + a_{22}r_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{21} & (a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{22} \\ (a_{21}r_{11} + a_{22}r_{21})a_{11} + (a_{21}r_{12} + a_{22}r_{22})a_{12} & (a_{21}r_{11} + a_{22}r_{21})a_{21} + (a_{21}r_{12} + a_{22}r_{22})a_{22} \end{bmatrix}$$

Therefore

$$\boldsymbol{\mathcal{T}}_{:,:,k} - \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^T =$$

$$\begin{bmatrix} t_{11} - ((a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{21}) & t_{12} - ((a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{22}) \\ t_{21} - ((a_{21}r_{11} + a_{22}r_{21})a_{11} + (a_{21}r_{12} + a_{22}r_{22})a_{12}) & t_{22} - ((a_{21}r_{11} + a_{22}r_{21})a_{21} + (a_{21}r_{12} + a_{22}r_{22})a_{22}) \end{bmatrix}$$

Now, we calculate $\boldsymbol{\mathcal{T}}_{:,:,k} - (\boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{\mathcal{R}}_{::k}$. First, recall

$$(\boldsymbol{A} \otimes \boldsymbol{A}) = \begin{bmatrix} a_{11}a_{11} & a_{11}a_{12} & a_{12}a_{11} & a_{12}a_{12} \\ a_{11}a_{21} & a_{11}a_{22} & a_{12}a_{21} & a_{12}a_{22} \\ a_{21}a_{11} & a_{21}a_{12} & a_{22}a_{11} & a_{22}a_{12} \\ a_{21}a_{21} & a_{21}a_{22} & a_{22}a_{21} & a_{22}a_{22} \end{bmatrix}$$

thus

$$(\boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{\mathcal{R}}_{::k} = \begin{bmatrix} a_{11}a_{11} & a_{11}a_{12} & a_{12}a_{11} & a_{12}a_{12} \\ a_{11}a_{21} & a_{11}a_{22} & a_{12}a_{21} & a_{12}a_{22} \\ a_{21}a_{11} & a_{21}a_{12} & a_{22}a_{11} & a_{22}a_{12} \\ a_{21}a_{21} & a_{21}a_{22} & a_{22}a_{21} & a_{22}a_{22} \end{bmatrix} \begin{bmatrix} r_{11} \\ r_{21} \\ r_{12} \\ r_{22} \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11}a_{11})r_{11} + (a_{11}a_{12})r_{21} + (a_{12}a_{11})r_{12} + (a_{12}a_{12})r_{22} \\ (a_{11}a_{21})r_{11} + (a_{11}a_{22})r_{21} + (a_{12}a_{21})r_{12} + (a_{12}a_{22})r_{22} \\ (a_{21}a_{11})r_{11} + (a_{21}a_{12})r_{21} + (a_{22}a_{11})r_{12} + (a_{22}a_{12})r_{22} \\ (a_{21}a_{21})r_{11} + (a_{21}a_{22})r_{21} + (a_{22}a_{21})r_{12} + (a_{22}a_{22})r_{22} \end{bmatrix}$$

$$\boldsymbol{\mathcal{T}}_{:,:,k} - (\boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{\mathcal{R}}_{::k} = \begin{bmatrix} t_{11} - ((a_{11}a_{11})r_{11} + (a_{11}a_{12})r_{21} + (a_{12}a_{11})r_{12} + (a_{12}a_{12})r_{22}) \\ t_{21} - ((a_{11}a_{21})r_{11} + (a_{11}a_{22})r_{21} + (a_{12}a_{21})r_{12} + (a_{12}a_{22})r_{22}) \\ t_{12} - ((a_{21}a_{11})r_{11} + (a_{21}a_{12})r_{21} + (a_{22}a_{11})r_{12} + (a_{22}a_{12})r_{22}) \\ t_{22} - ((a_{21}a_{21})r_{11} + (a_{21}a_{22})r_{21} + (a_{22}a_{21})r_{12} + (a_{22}a_{22})r_{22}) \end{bmatrix}$$

Taking the norm of both expressions leads to

$$||\boldsymbol{\mathcal{T}}_{:,:,k} - \boldsymbol{A}\boldsymbol{\mathcal{R}}_{:,:,k}\boldsymbol{A}^{\mathrm{T}}||^2 = ||\mathrm{vec}(\boldsymbol{\mathcal{T}}_{:,:,k}) - (\boldsymbol{A} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{\mathcal{R}}_{:,:,k})||^2 \ .$$

2. **Stochastic Gradient Descent (SGD)**

Recall that the score function of RESCAL is given by

$$f^{RESCAL}((e_i, r_j, e_k)) = \boldsymbol{a}_i \boldsymbol{\mathcal{R}}_{:,:,k} \boldsymbol{a}_j^T = \sum_{b=1}^{l} \sum_{c=1}^{l} \mathcal{R}_{bck} a_{ib} a_{jc} \ .$$

The partial derivatives with respect to its parameters (i.e. $\mathcal{R}_{bck}, a_{ib}$, and a_{jc}) are given by

$\frac{\partial f^{RESCAL}((e_i,r_j,e_k))}{\partial \mathcal{R}_{bck}} = a_{ib}a_{jc}$

$\frac{\partial f^{RESCAL}((e_i,r_j,e_k))}{\partial a_{ib}} = \sum_{c=1}^{l} \mathcal{R}_{bck} a_{jc}$

$\frac{\partial f^{RESCAL}((e_i,r_j,e_k))}{\partial a_{jc}} = \sum_{b=1}^{l} \mathcal{R}_{bck} a_{ib}$.