

Knowledge Graph Analysis

Solutions of Exercise Sheet 5

Dr. Asja Fischer, Prof. Jens Lehmann

December 27, 2016

1 IN CLASS

1a) Alternating Least Squares (ALS)

We want to show that¹ $\hat{\mathcal{T}} = \mathcal{R} \times_1 \mathbf{A} \times_2 \mathbf{A}$ is equivalent to

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

Using the definition of the *Kronecker Product* one gets

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

and thus

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

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

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

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

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

1b) To show that the equivalence

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

holds for

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

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

$$\begin{aligned}\mathbf{A}\mathcal{R}_{::,k}\mathbf{A}^T &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}r_{11} + a_{12}r_{21} & a_{11}r_{12} + a_{12}r_{22} \\ a_{21}r_{11} + a_{22}r_{21} & a_{21}r_{12} + a_{22}r_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \\ &= \begin{bmatrix} (a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{21} & (a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{22} \\ (a_{21}r_{11} + a_{22}r_{21})a_{11} + (a_{21}r_{12} + a_{22}r_{22})a_{12} & (a_{21}r_{11} + a_{22}r_{21})a_{21} + (a_{21}r_{12} + a_{22}r_{22})a_{22} \end{bmatrix}\end{aligned}$$

Therefore

$$\begin{aligned}\mathcal{T}_{::,k} - \mathbf{A}\mathcal{R}_{::,k}\mathbf{A}^T &= \\ &= \begin{bmatrix} t_{11} - ((a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{21}) & t_{12} - ((a_{11}r_{11} + a_{12}r_{21})a_{11} + (a_{11}r_{12} + a_{12}r_{22})a_{22}) \\ t_{21} - ((a_{21}r_{11} + a_{22}r_{21})a_{11} + (a_{21}r_{12} + a_{22}r_{22})a_{12}) & t_{22} - ((a_{21}r_{11} + a_{22}r_{21})a_{21} + (a_{21}r_{12} + a_{22}r_{22})a_{22}) \end{bmatrix}\end{aligned}$$

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

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

thus

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

Taking the norm of both expressions leads to

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

2. Stochastic Gradient Descent (SGD)

Recall that the score function of RESCAL is given by

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

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

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