

---

# Knowledge Graph Analysis

## Exercise Sheet 5

---

Dr. Asja Fischer, Prof. Jens Lehmann

December 4, 2017

### 1 IN CLASS

#### 1. Alternating Least Squares (ALS)

a) In the first step of ALS (slide 16 of the lecture) we made use of the equivalence

$$\begin{aligned} \|\mathcal{T} - \mathcal{R} \times_1 \mathbf{A}_l \times_2 \mathbf{A}_r\|^2 &= \|\mathcal{T}_{(1)} - \mathbf{A}_l \mathcal{R}_{(1)} (\mathbf{I} \otimes \mathbf{A}_r)^T\|^2 \\ &= \|\mathcal{T}_{(2)} - \mathbf{A}_r \mathcal{R}_{(2)} (\mathbf{I} \otimes \mathbf{A}_l)^T\|^2 \end{aligned}$$

To understand this show that  $\hat{\mathcal{T}} = \mathbf{R} \times_1 \mathbf{A} \times_2 \mathbf{A}$  is equivalent to

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

1

Recall the definition of the *Kronecker Product*:

Let  $A \in \mathbb{R}^{Q \times R}$ ,  $B \in \mathbb{R}^{S \times T}$ . The Kronecker Product is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11} \mathbf{B} & \dots & a_{1n} \mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1} \mathbf{B} & \dots & a_{mn} \mathbf{B} \end{bmatrix}$$

where  $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{QS \times RT}$ .

---

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

- b) In the second step of ALS (slides 17+18 of the lecture) we made use of the equivalence

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

where  $\text{vec}(M)$  is the vector which results from stacking the columns of the matrix  $M$ .

Show that this equivalence is true for

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

## 2. Stochastic Gradient Descent (SGD)

Recall that the score function of RESCAL is given by

$$f^{RESCAL}((e_i, r_j, e_k)) = \mathbf{a}_i \mathcal{R}_{::,k} \mathbf{a}_j^T = \sum_{a=1}^l \sum_{a=1}^l \mathcal{R}_{abk} a_{ia} a_{jb} .$$

Calculate the gradient  $\nabla f^{RESCAL}((e_i, r_j, e_k))$  by specifying the partial derivatives with respect to its parameters (i.e.  $\mathcal{R}_{abk}$ ,  $a_{ia}$ , and  $a_{jb}$ ).