COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF BONN

Knowledge Graph Analysis

Exercise Sheet 5

Dr. Asja Fischer, Prof. Jens Lehmann

December 27, 2016

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

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

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

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

   [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

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

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

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

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

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

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

Show that this is equivalence is true for

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

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

Recall that the score function of RESCAL is given by

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

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}).