D5: Databases and Information Systems
Tensors in Data Analysis, WS 2017–18
Homework #3: Variations and applications of the CP decomposition
Tutorial: **12 October 2017** at 14:30

You can discuss these problems with other students, but everybody must do and present their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You are of course free to use material from the Internet, but again, you must present the intermediate steps and you must also be able to explain why the steps are valid and why you chose them. You can mark an answer even if it is not complete or correct, as long as you have made significant progress towards solving it. Note, however, that the lecturer does the final decision on whether your solution is complete (or correct) enough for a mark.

Problem 1 (Maximum rank). It was stated in the lectures that the rank of a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I \times J \times K}$ is never more than

$$\min\{IJ, IK, JK\} .$$

Let I, J, and K be that $JK = \min\{IJ, IK, JK\}$ and let $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I \times J \times K}$ be arbitrary. Your task is to construct $\boldsymbol{A} \in \mathbb{R}^{I \times JK}$, $\boldsymbol{B} \in \mathbb{R}^{J \times JK}$, and $\boldsymbol{C} \in \mathbb{R}^{K \times JK}$ such that

$$\boldsymbol{T}_{(1)} = \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T .$$

Hint: Construct $\boldsymbol{B}$ from identity matrices.

Problem 2 (Nonnegative INDSCAL). Present an algorithm for nonnegative 3-way INDSCAL. That is, given a nonnegative 3-way tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}_{\geqslant 0}^{I \times J \times K}$ and an integer R, find matrices $\boldsymbol{A} \in \mathbb{R}_{\geqslant 0}^{I \times R}$, $\boldsymbol{B} \in \mathbb{R}_{\geqslant 0}^{J \times R}$, and $\boldsymbol{C} \in \mathbb{R}_{\geqslant 0}^{K \times R}$ that aim at minimizing

$$\|\boldsymbol{\mathcal{T}} - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]\| .$$

Problem 3 (CP-APR for KL-divergence). In CP-APR, we need to find a matrix $\boldsymbol{A}$ that minimizes

$$L(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T - \boldsymbol{T}_{(1)} * \log(\boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T) .$$

This is a type of a KL divergence. In nonnegative matrix factorization (NMF), we are given a nonnegative matrix $\boldsymbol{A} \in \mathbb{R}_{\geqslant 0}^{I \times J}$ and an integer K and we have to find nonnegative matrices $\boldsymbol{W} \in \mathbb{R}_{\geqslant 0}^{I \times K}$ and $\boldsymbol{H} \in \mathbb{R}_{\geqslant 0}^{K \times J}$ such that $\boldsymbol{A} \approx \boldsymbol{W}\boldsymbol{H}$.

The standard NMF algorithm for KL divergence has the following update rule:

$$\boldsymbol{W}_{ik} \leftarrow \boldsymbol{W}_{ik} \frac{\sum_{j=1}^{m} (\boldsymbol{A}_{ij}/(\boldsymbol{W}\boldsymbol{H})_{ij}) \boldsymbol{H}_{kj}}{\sum_{j=1}^{m} \boldsymbol{H}_{kj}} .$$

Adapt this update rule for the factor matrix $\boldsymbol{A}$ in the CP decomposition. How does it relate to the update rule

$$\boldsymbol{A} \leftarrow \boldsymbol{A} * \big(\boldsymbol{T}_{(1)} \oslash (\boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T)\big)(\boldsymbol{C} \odot \boldsymbol{B})^T ,$$

presented in the lecture? (To recap, $\oslash$ is the element-wise division.)

Problem 4 (PARAFAC2). The PARAFAC2 decomposition is another variant of the CP decomposition, defined slice-wise as follows. Given K matrices $\boldsymbol{X}_k \in \mathbb{R}^{I_k \times J}$ and rank R, find K matrices $\boldsymbol{U}_k \in \mathbb{R}^{I_k \times R}$, diagonal matrices $\boldsymbol{S}_k \in \mathbb{R}^{R \times R}$, and a matrix $\boldsymbol{V} \in \mathbb{R}^{J \times R}$ such that

$$\sum_{k=1}^{K} \left\| \boldsymbol{X}_k - \boldsymbol{U}_k \boldsymbol{S}_k \boldsymbol{V}^T \right\|_F$$

is minimized.

a) PARAFAC2 is related to CP, but how? Under which conditions is PARAFAC2 the same as the CP decomposition?

b) Consider following kind of health records data: We have longitudinal health records data over K patients and J attributes, such as diagnoses and medication. For each patient, we have collected these attributes over different time span and at different times, and each patient k is represented by a I_k-by-J matrix $\boldsymbol{X}_k$, where I_k is the number of observations for this patient, and $(\boldsymbol{X}_K)_{ij}$ is the value of variable j that observation point i. Notice that the observation points do not align between the users, that is, they correspond to different points in time. Assume we do rank-R PARAFAC2 to the collection of such matrices $\{\boldsymbol{X}_k\}_{k=1}^K$ and obtain $\{\boldsymbol{U}_k, \boldsymbol{S}_k\}_{k=1}^K$, and $\boldsymbol{V}$.

   We can assume that the columns of the J-by-R matrix $\boldsymbol{V}$ corresponds to some latent *phenotypes*, that is, they encode which diagnoses and medication "go together." How would you interpret the other factors?