Homework 2, CS 6241 Spring 2020

Instructor: Austin R. Benson

Due Thursday, March 5 2020 at 11:59pm ET on CMS

Theoretical analysis

1. Rank-1 NMF

We discussed in class how, in general, finding a best rank-k NMF approximation is NP-hard. Here, we analyze a case that we can solve. Let A be an $m \times n$ nonnegative matrix. Consider a rank-1 NMF approximation:

$$\min_{x \in \mathbb{R}^m,\, y \in \mathbb{R}^n} \quad \|A - xy^T\|_F^2 \qquad \text{subject to } \; x, y \geq 0 \tag{1}$$

Find a minimizer for eq. (1) in terms of a truncated SVD of A.[1]

[1] Hint: Use the Perron-Frobenius theorem on AA^T and A^TA.

2. Symmetric rank-1 matrix and tensor approximation

(a) Let A be an $n \times n$ symmetric matrix. The best symmetric rank-1 approximation problem can be written as

$$\min_{\gamma \in \mathbb{R},\, x \in \mathbb{R}^n} \quad \|A - \gamma \cdot xx^T\|_F^2 \qquad \text{subject to } \; \|x\|_2^2 = 1. \tag{2}$$

Let $L(\gamma, x, \lambda)$ be the Lagrangian for eq. (2). Show that at any stationary point $(\gamma^*, x^*, \lambda^*)$, x^* is a unit-norm eigenvector of A with eigenvalue γ^*. Using this, what is the minimizer for eq. (2)?

(b) Now let $\underline{A}$ be an $n \times n \times n$ symmetric tensor, where "symmetric" means that the entries of $\underline{A}$ remain the same under any permutation of indices:

$$\underline{A}_{ijk} = \underline{A}_{ikj} = \underline{A}_{jik} = \underline{A}_{jki} = \underline{A}_{kij} = \underline{A}_{kji}. \tag{3}$$

A *tensor Z-eigenpair* is a vector x and scalar λ satisfying

$$\sum_{1 \leq j,k \leq n} \underline{A}_{ijk} x_j x_k = \lambda x_i, \; i = 1, \ldots, n, \qquad \|x\|_2^2 = 1, \tag{4}$$

We call x a *tensor Z-eigenvector* and λ a *tensor Z-eigenvalue*. Unlike matrix eigenvectors, Z-eigenvectors are not scale-invariant, so the size constraint on x is necessary for the definition of a Z-eigenvalue here. Let's assume that $\underline{A}$ has finitely many Z-eigenpairs.[2] Show that a minimizer for

$$\min_{\gamma \in \mathbb{R},\, x \in \mathbb{R}^n} \quad \|\underline{A} - \gamma \cdot x \otimes x \otimes x\|_F^2 \qquad \text{subject to } \; \|x\|_2^2 = 1 \tag{5}$$

corresponds to a tensor Z-eigenpair.[3]

[2] This is usually the case. See *The number of eigenvalues of a tensor*, Cartwright and Sturmfels, LAA, 2013.

[3] Unlike what we saw in class for rank-2 approximation, a best rank-1 approximation always exists. However, computing the best rank-1 approximation (symmetric or not) is still NP-hard; see *Most tensor problems are NP-hard*, Hillar and Lim, JACM, 2013.

Data analysis

By the time this assignment is due, we will have covered a number of techniques for dimensionality reduction and latent factor models:

1. PCA and robust PCA
2. Matrix Completion
3. Nonnegative matrix factorization
4. The Interpolative decomposition and CUR
5. Tensor decompositions
6. Non-linear dimensionality reduction

Implement one method from class (or a related method). Your implementation should not just call a library function of that method. Instead, you should implement the basic computations (for this, you can use libraries). Many of the Jupyter notebooks from class follow this pattern.[4] You are permitted to re-use code from the course Jupyter notebooks for this part of the assignment. Include a brief description of the method and why it can be useful.

Next, find a "real-world" (i.e., not synthetic) dataset[5] on which to apply the method you implemented. Provide a brief description of the dataset and explain what the numerical method reveals about the data. Provide some qualitative insights.

[4] `https://github.com/arbenson/cs6241_2020sp`

[5] The course web site (`https://www.cs.cornell.edu/courses/cs6241/2020sp/`) has some pointers for datasets. You can also use datasets from your own work / research.

Preparation & submission guidelines

Typesetting. All homeworks should be prepared with LaTeX. Handwritten homeworks will not be accepted.

Code. Part of the assignment involves writing code. You need to include your code in your submission, and you can easily do so using the `listings` package. You do not need to include code that you write for the qualitative analysis.

Collaboration. You are encouraged to discuss and collaborate on the homework. However, you have to write up your own solutions and write your own code. You must also list your collaborators on your homework.

Academic Integrity. We expect you to maintain academic integrity in the course. For example, follow the collaboration guidelines above and do not just copy someone else's code. Failure to maintain academic integrity will be penalized severely. Plagiarism is a form of academic misconduct, so make sure to provide proper citations. Cornell has a number of guidelines on plagiarism.[6]

[6] `https://plagiarism.arts.cornell.edu/tutorial/index.cfm`

Submission. Your homework should be submitted as a single PDF that includes your solutions to the theoretical analysis and data analysis components, along with your code (as outlined above). Also include your name and the names of any collaborators. Submit your PDF on CMS.[7]

[7] `https://cmsx.cs.cornell.edu`