

CmpE 482 - Homework 1 (Due: October 16th, 2017 midnight)

Description

This exercise is about using the Singular Value Decomposition for the temporal link prediction problem. When you load *hw1_data.mat*, you will see two data sets, i.e., \mathbf{X} and \mathbf{Y} . \mathbf{X} shows the number of papers published by authors at various conferences between 1991 and 2004. It is of size 471 (authors) \times 366 (conferences) \times 14 (years). Given \mathbf{X} , we want to predict who is going to publish at which conference in 2005. Matrix \mathbf{Y} shows the ground truth, i.e., publications in 2005.

Execute the following steps and return the outputs in the deliverables:

- Change each nonzero entry of \mathbf{X} as $x_{ijk} = \log(x_{ijk}) + 1$, where $x_{ijk} \neq 0$.
- Collapse the three-way array \mathbf{X} by summing up over the years mode and form an authors by conferences matrix of size 471×366 . Let this matrix be \mathbf{Z} .
- Compute the SVD of \mathbf{Z} .
- Construct the best rank- K approximation of \mathbf{Z} denoted as $\hat{\mathbf{Z}}_K$ for different K values, i.e., $K = \{2, 10, 20, 50, 100, 300\}$. Entries of $\hat{\mathbf{Z}}_K$, i.e., $\hat{\mathbf{Z}}_K(i, j)$ can be used as scores to predict if there is a link between the i^{th} author and j^{th} conference in 2005. A link means an author publishes at a conference.
- Replace every nonzero entry of \mathbf{Y} with 1. Vectorize \mathbf{Y} , i.e., $\mathbf{Y}(:)$ in MATLAB notation, which will correspond to the true labels (0's and 1's).
- For each value of K , vectorize $\hat{\mathbf{Z}}_K$, which will correspond to the scores/predictions.
- For each value of K , plot the Receiver Operating Characteristics (ROC) Curve and calculate the area under the curve (AUC) (Note: You can use the `perfcurve` function in MATLAB).

Deliverables

Work in pairs and send the following by email to evrim.acarataman@gmail.com:

- Your MATLAB script as an .m file
- Plot of singular values of matrix \mathbf{Z}
- The ROC plots (preferably as a single figure with all the curves for different K values) where AUC values are shown somewhere in the figure.

Bonus points

Any modification that gives an higher AUC value than the ones you get by following the instructions will get extra points.