Assignment 2

Matrix Computations and Applications
1 Practical problems

This part of the assignment should be solved individually. Hand in a printed copy of your source code including a path to the digital version. A brief report (about 3 A4 sheets in addition to the front page) is required for this part of the assignment. It should include convergence plots and discussions of the results. You should be prepared to explain the source code and the results in detail.

1.1 Query matching

Write a MATLAB function that performs query matching in the vector space model using the modified cosine formula, as explained in lecture C5 and reference [1]. Let the user input a term-by-document matrix $A$, a representation rank $k$, and a low-rank approximation method (either QRP or SVD). Test the function on the two toy collections given on the course web.

1.2 Non-negative matrix factorization

Write a MATLAB function that implements the multiplicative method for NMF explained in lecture C5 and reference [2]. Exploit the MATLAB notation (e.g., .* and ./) to implement the point-wise formulas efficiently.

Write a MATLAB function that implements the alternating least squares method for NMF explained in lecture C5 and reference [3]. Solve each unconstrained linear least squares problem using a QR factorization of the coefficient matrix. You may assume that the coefficient matrix is not rank-deficient.

1.3 Topic extraction

Using the two toy collections given on the course web (see below), extract topics via NMF for the representation ranks $k = 2$ and $k = 3$. Then use the NMF to cluster the documents into $k$ clusters. Show convergence plots (for both NMF algorithms) and discuss the quality of the resulting document clusters. Do the two NMF algorithms give you similar topics and clusters? Why/why not?

Example collections

- babies.m (http://www8.cs.umu.se/kurser/5DA002/HT11/babies.m)
- recipes.m (http://www8.cs.umu.se/kurser/5DA002/HT11/recipes.m)

References

