Linear Algebra Methods for Data Mining

Saara Hyvönen, Saara.Hyvonen@cs.helsinki.fi

Spring 2007

Course details
Lectures

Lecture time: Tue 12-14, Wed 14-16 in C220.

Lecturer: Dr. Saara Hyvönen, Saara.Hyvonen@cs.helsinki.fi

Office hours: Tue 14-15 in room A347

Course language: English

Exercise sessions:

Time: Fri 12-14 BK106

Teacher: Juan Carlos Borrás García, borras@cs.helsinki.fi

Office hours: Wed 13-14 in room A309

Language: English

What do you need to do to get the credits?

- Both:
  - 5 sets of homeworks, to be handed in
  - Exam Friday March 2nd at 16-19

- Homeworks and exam are of equal weight.

- The homeworks will be available on the course web page every Friday. They are due the next Friday, except for two larger project works, which will be due in 2 weeks.
Course material

• Lecture notes will be available on the web page after each lecture.

• The course follows to some extent the book by Lars Eldén: Matrix Methods in Data Mining and Pattern Recognition, SIAM 2007. A copy of the draft can be obtained directly from lecturer.

• References to other, potentially useful material can be found on the course web page.
Course outline, version 1:

- Preliminaries: basic linear algebra
- SVD and related methods (PCA, LSI)
- Other matrix decompositions (eigenvalue decomposition, QR, NMF)
- LDA (Linear Discriminant Analysis)
- Finding eigenvalues and eigenvectors using the power method
Course outline, version 2:

• Term-document matrices
• Recognizing hand-written digits (SVD)
• Text mining and information retrieval (LSI)
• Analyzing spatial data (PCA)
• Page Ranking for a Web Search Engine (power method)
• Classifying tissue samples based on gene expression values (LDA)

Attn: Reliability of outlines questionable! Changes may occur!