Open your mind. LUT.
Lappeenranta University of Technology
LUT School of Engineering Science

2015-12-9

BM20A6100 Advanced Data Analysis and Machine Learning

Heikki Haario and Lasse Lensu

Practical Assignment

Atmospheric Data Analysis

1 Problem Statement

Your task is to analyse atmospheric data by using computational data analysis tools. The data to be used is given as background information. The general problem in the analysis is to study data analysis and dimensionality reduction from the following viewpoints:

- Variables explaining the variation in and dynamics of the data.
- Retaining relevant characteristics of the data.

Your task is to implement the data analysis and visualisation in Matlab. For approprate focusing of the task, you have to select one topic from the following list and enrol to the corresponding group in Moodle (also your group partner should do this).

1.1 Topic 1: Standard analysis and animations of spatiotemporal data

Experiment with the spatiotemporal data by using the standard data analysis method, that is, principal component analysis (PCA) / singular value decomposition (SVD). Study the clusters in the data to find out variables explaining the time- and location-dependent (2-D) variation.

From the provided spatial data, utilise a single 2-D layer at a time. Produce animations visualising the dynamics of the temperature and pressure changes with time.

1.2 Topic 2: Random projections of spatiotemporal data

Experiment with the spatiotemporal data by using random projections (RPs). Study the data to find out variables explaining the time- and location-dependent (3-D) variation.

From the provided spatial data, utilise the full 3-D location information. Visualise the dynamics of the temperature and pressure changes with time.

1.3 Topic 3: Random projections and retaining relevant data characteristics

Experiment with the spatiotemporal data by using RPs from the viewpoint of retaining relevant characteristics of the data. Study the Johnson-Lindenstrauss lemma, and experimentally test how it holds. Try to answer to the following questions: Which characteristics of the data are preserved? In what sense? You might use your own synthetic data as well.

PO BOX 20, FI-53851 LAPPEENRANTA, FINLAND
E-mail: lasse.lensu@lut.fi

1.4 The data

The data to be used consists of temperature and pressure readings at different locations of the world and at different timepoints. The locations form either a 2-D or 3-D grid, and the timepoints are averages over a time window.

The data for the experimentation is provided through an Internet server. This is because of the file size is large and there are limitations in Moodle. The data can be downloaded from air.mon.mean.nc

2 Requirements

The practical assignment is meant to be done in freely selected groups of two persons. If someone wants to work alone, this is also possible.

To carry out the *programming task*, you must obey the following rules:

- Allowed: Use of standard and external Matlab toolboxes, and other software if it can be used from Matlab.
- Not allowed: Use of any code prepared by someone else without properly mentioning the source(s).

To prepare the *documentation* of your work, you must obey the following rules:

- Allowed: Use of references if you acknowledge them (proper citation to the reference used) and do not copy text from a reference.
- Not allowed: Use of any material prepared by others (without properly acknowledging the source), or direct copying of sentences or their parts from a reference.

By returning the assignment you assure that i) you acknowledge all sources (no plagiarism), and ii) you have not used any forbidden material.

2.1 Implementation

For the implementation, use primarily standard Matlab and its toolboxes. Also other software can be used as long as it is programmed in Matlab or is callable from Matlab. In this case, you must acknowledge the source of the software.

Remember to properly comment your codes. Write also a help section to your codes that tells the purpose of the function, usage, and explanation of the parameters.

2.2 Documentation

Write a report in English about your project. The documentation should include a cover page where you give the course number and name, project title, date, and the names and student numbers of the authors.

Describe the methods used for data analysis in such detail that a reader understands your approach and would be able to reproduce your results. Cite the relevant references for enabling the understanding of the approach. Justify your choices by presenting the grounds to select the methods for your solution.

Include in the report the main results of your experimentation. Analyse the results critically.

3 Deadline and submission

The deadline of submitting the results of your work to Moodle is **Sunday, 31 January 2016 at 23:55 EET**. The results containing the documentation in pdf format and all relevant codes must be packed into a single file using `zip`, or `tar` and `gzip`. The file name of the package must be `stnum.zip` or `stnum.tar.gz` where `stnum` is the student number of one of the authors. When `stnum.zip` is extracted, it should create a single directory `stnum`. This directory should contain a `readme.txt` file instructing how to start using your implementation.

4 Grading

The work will be evaluated based on the submitted report and implementation according to the following criteria:

1. Methods: The documentation contains grounds for method selections and such descriptions that enable understanding the solution.
2. Successful implementation: The code is properly commented. Based on the results presented in the report, the implementation works for the described purpose.
3. Results and analysis: The produced results are appropriately presented and analysed in depth to support the conclusions in the report.
4. Documentation: The report includes all the standard parts and relevant references.

In addition to the "standard" criteria described above, the following criteria will be taken into account in the evaluation:

- Novelty of the approach or results. (If the idea comes from somebody elses work, you must acknowledge the source.)
- The difficulty of implementing the selected approach. (However, the result of the work cannot be empty.)
- Well-designed demonstration or visualisation.

Based on the evaluation, the work will be graded using the standard scale 0, 1,..., 5. the practical assignment grade affects the final grade of the course, and to pass the course, the grade of the practical assignment must be at least 1.

5 Notes and tips

If there are any problems with the assignment description and/or data, contact the person supervising the practical assignment. This should be done before inventing your own interpretations or making radical assumptions.