Machine Learning

Lab 1: Introduction

1 Preamble

Site license UCLA has a campus-wide license for MATLAB:

see https://softwarecentral.ucla.edu/matlab-getmatlab for more information/instructions.

If you do not want to install MATLAB, locally (it is quite big), you may use "Apporto"-virtual machines (be in touch with support@pic.ucla.edu for help on that). Most conveniently, though, we suggest using MATLAB on the web:

MATLAB on the web To avoid the need of installation, we suggest using MATLAB on the web, instead. This service is available to students, staff, and faculty of UCLA at no cost. The only requirement is to sign up for a MATLAB account with a ucla.edu-email-address.

Follow the instructions on https://www.mathworks.com/academia/tah-portal/ucla-31454052.html to create your UCLA-MATLAB account. Once complete, you have access to MATLAB online at matlab.mathworks.com (iOS safari browser is discouraged). Note that your MATLAB workspace is persistent across sessions, devices, and there is an online filesystem!

MATLAB Reports Your solutions to the lab assignments should include a clear write up of your solutions (LaTeX is strongly preferred, but other text editors or handwritten form is fine as well) including an English description of your solution to the programming exercises (explain your process and the high level action of your code). Reports should include answers to the specific questions/tasks set forth in the assignment. Please also include a discussion and reflection of what the practical experiments reveal, to you. For each MATLAB assignment, you will need to submit a single .pdf-file to gradescope with your code, experiments, figures, results, and observations.

.m-files or Livescripts You are encouraged to organize your online filesystem appropriately. It is recommended that for each MATLAB assignment, you create one or more .m-files or livescripts as required. Make sure to include an appropriate amount of comments to make your code understandable. Use disp(...) to print values of variables on the screen, and use plots and figures as suitable. You can create sections to your script (for reporting purposes) by using %%-signs. Use comments at the end of the file to discuss results and observations as required.

publish and submit A simple way is to use MATLAB's publish-function to create .pdf-reports. The publish button will run the current file, and render the code along with output (command line and figures) into a single .pdf-file. This method is simple enough—it is your responsibility though to make sure the report is complete! Note: The publish function does in general not play well with functions that require parameters, and scripts that include user interaction (e.g. input). Use a wrapper script that calls such a function with parameters, and include the code of the inner function using <include>.

A few helpful items¹:

- Use %% ... to create a new section; this embeds all results of a section before moving on to the next.
- Use % comments... to include your observations.
- To include the code of *separate* function files, use <include> *exactly* as follows (extra comment lines and spacing between % and <, etc., matters!):

¹ Check https://www.mathworks.com/help/matlab/matlab_prog/marking-up-matlab-comments-for-publishing.html

```
%% Beginning of the wrapper script which we will publish
%
% <include>SeparateFunctionFileToBeIncludedInTheReport.m</include>
%
```

LATEX reports A possibly preferable alternative to MATLAB's publish feature is to use LATEX to properly typeset your report, including background, comments, code, experimental setup, results, discussion, and conclusion.

MATLAB due dates Either way, submit your report as a single pdf-file per assignment, through gradescope. Your reports for the 8 labs will be due on Fridays, each week (see schedule, below), by 6:00pm PST.

2 Overview

In this lab we will review the basic functionality of MATLAB to load, visualize, and perform basic operations on a data set. These lab assignments are organized as follows: the overview provides a simple summary of what will be done in the lab, the next section provides a list of provided resources which come with the lab, the guide is intended to give you a step-by-step procedure to arriving at lab goals, and finally a section at the end provides extra questions which are not graded but provide additional questions which build on the conclusions of the lab assignment. Each assignment is graded out of twenty points. The "Provided Resources" section contains an itemized list of materials that have been given to you. Resources ending in .mat contain data for MATLAB, and can be loaded using either the MATLAB function load() or by dragging the file into the MATLAB window. Resources ending in .m contain MATLAB code, either as a function that is completed and provided or as a placeholder for code which you have to complete. Each of these files begin with a comment section which describe precisely what the code either does, or is intended to do. For example, be aware of the orientation of data for each function: you might be passing M when the code assumes it was passed M' (the transpose).

Once you have a basic understanding of the resources provided, it is on to the "Guide" section. Here the lab is outlined in steps – simply start from the beginning and work your way to the end. Some steps may simply ask you to look at something for your own edification, while others will be graded. If a step of the lab is graded then there will be a number in parenthesis indicating the point value.

Generally graded steps fall into three groups: completed code for algorithms which are used in the lab, figures/plots generated from the data you encounter, or written responses to questions about the data. Responses should be only a couple sentences in length and will be graded on the quality of thoughtfulness. If a question is asked in the guide, it is implied that an answer will be turned in. Everything else that must be submitted will be explicitly stated as such.

The "Extra" section can be completely ignored, if you like. The questions here serve to extend the lab in interesting ways that rest beyond the original scope.

Notice: We will sometimes be working with large collections of data, which means that running code might take a very long time to finish. Try to write your code as efficiently as possible. If something is taking more than a minute to run with no end in sight, abort execution (in the command window press CTRL+c) and reduce the amount of data. Accurate results obtained with less than the full collection of data will be accepted.

3 Provided Resource

data.mat - A single matrix X with one observation per column. The data was simply generated at random.

4 Guide

1. When using large collections of data, efficient programming is often important. Perform a brief experiment by comparing the time MATLAB takes to run through a loop versus the time the sum command requires. Do this by loading the matrix *X* from data.mat, then implementing the algorithm:

```
s <- 0
for i = 1 to 1,000,000
    s <- s + |X(:,i)|
end for</pre>
```

which calculates the sum of the column ℓ_2 -norms of X (that is, Euclidean distance). The result should be $s\approx 1.4\cdot 10^6$. Now, perform the same calculation using the function sum with no for statement and verify that the number you get is the same.

2. Time the result of the two approaches above and turn in the time required for each method. You can time a piece of code by calling tic before the code to be timed, followed by too which evaluates to the time, in seconds, since tic was called.

5 Extra

None for this lab.