Course Policies Handout

Contents

1 From the Bulletin:
2 Structure of the course:
3 How to get the most out of the labs
4 Required software and books:
5 Collaboration policy:
6 Asking questions:
7 Grading:

From the Bulletin:

The interpretation of data is key to making new discoveries, making optimal decisions, and designing experiments. Students will learn skills of data analysis through hands-on, computer-based tutorials and exercises that include experimental data from the biological sciences. Knowledge of very basic statistics (mean, median) will be assumed. Usually offered every second year.

Structure of the course:

In-class Labs – Classroom time is devoted to self-guided tutorials. The professor may occasionally give introductory remarks, and will be available for questions.

Problem sets / homework sets – Homework sets are completed outside of class on the computer. There are 6 homework sets during the course.

Team projects – At about 3 points during the course, there will be a team exercise. Students will split into teams to create 1 or 2 figures that address a particular question. One member of the team will briefly present the figure during a designated class session.

The course is designed in a hands-on manner because it is crucial to integrate the conceptual thinking with the mathematical and computer skills needed to actually carry out the analysis. The in-class labs are designed to break down, into very small parts, the concepts and skills that are required to address certain questions. The purpose of the homework is to allow students to practice what they have learned and to develop the ability to solve similar problems independently. The team projects allow students to work in groups, much as one would in a real research laboratory, in order to share ideas, solutions, and criticisms, and to receive feedback from other groups.

How to get the most out of the labs

Some of you may be doing these labs for credit in the course, others may be doing them to learn how to use Matlab on your own or as you are working in a lab. To get the most out of the labs:

- Type in the example code yourself, rather than cutting and pasting. If you type it out yourself, you'll start to develop familiarity with the language.
- Make sure you go slowly enough for the material to sink in. We all get used to skimming web pages, but you'll get more out of the tutorials if you can slow yourself down and think as you read.
- Even if you are not taking the course for credit, answer the questions as you go. This will help to ensure that you are thinking about the material as you go.

Required software and books:

The course requires access to Matlab. Matlab is available in the Goldfarb and Farber computer clusters here on campus, and is also available in many research laboratories. Students who don't have easy access to Matlab in one of these locations can install Matlab using the university's site license; see Setting Up Shop at Home.

In addition, I've required 2 textbooks; 1 is a basic introduction to Matlab for scientists and engineers, while the other is a statistics book for the biological sciences. While I believe the tutorials will make it easy to learn the concepts and to work with data, it will be useful to have these textbooks to use as a reference. These books are on reserve in the Main Library on campus. At various times, you will need access to the books to complete some of the problem sets.


In the 2011-2012 course, we used this Matlab text that may be of interest to some students:


Collaboration policy:

Students are encouraged to discuss the homework sets and help each other, with some limitations:

- Students should attempt each homework question for 20 minutes before seeking help on that portion from other students or the professor.
- Students may talk about concepts, discuss lines or snippets of code in emails, message boards, or on a whiteboard, and may debate whether various solutions are correct.
- Students are not permitted to exchange complete `.m` files; each student should write their own `.m` files themselves and understand what they are turning in.
- When you have consulted other students while completing your homework, credit them by including their names.

Asking questions:
We will use the site piazza.com to handle questions. When anyone asks a question, the entire class will be able to see the answer. If the question is not a clarifying question, I am very likely to give a hint rather than spelling out the solution, so leave yourself plenty of time to complete the assignments. See the "Ask a question" link on the left side bar to log-in. You can also send a private question to the instructor if you have a question about your grade.

Note: Please do not send course questions to bio107asubmit <at> gmail.com. I only use this account for grading, and I don't check it frequently.

In addition, there are office hours listed on the Instruction Team page.

Grading:

The overall course grade will consist of a weighted average of scores for in-class labs (20%), out-of-class problem sets (due about once per 2 weeks) (45%), and team projects (35%). There is no final exam but problem sets will probably require more time for completion than in a typical course. **It is important to begin work on the problem sets early, as typically they will be difficult to complete on the last day.** Each student can turn in 1 assignment up to 4 days late without penalty; other problem sets will lose 10% of the maximum credit per hour they are late.

**4 credit hour effort statement:**

Success in this 4 credit hour course is based on the expectation that students will spend a minimum of 9 hours of outside time per week.
Topics covered

Units of the course:

1. **Did my drug work?**
   1. Basic use of Matlab, including variables, functions
   2. Plotting (bar plots, scatter plots, histograms, cumulative histograms)
   3. Statistical inference
   4. Comparing distributions (Kolmogorov-Smirnov test), means (T-test)
   5. ANOVA

2. **Fits and indexes**
   1. How to create a fit (linear fits, non-linear fits)
   2. Dealing with a mountain of data while avoiding garbage in / garbage out
   3. Calculating robustness/uncertainty for fits and index values

3. **Time series**
   1. Correlations in time
   2. Feature detection
   3. Fourier analysis / frequency analysis
   4. Convolution and filtering

4. **Image processing**
   1. Digital representations of images in the world
   2. Feature detection and extraction
   3. Functional imaging (time series of images)

5. **High-dimensional data**
   1. Clustering
   2. Dimensionality reduction (including principle component analysis)
Course Calendar

Note: all lab responses are due 1 week after the lab is done in class.

Class meets 5pm - 6:20pm

August/September 2015

8/31(M)  Lab 1.1 Getting started: Matlab and Data Analysis/Stats basics
9/02(W)  Lab 1.2 Variables, Functions, and Simple Plots
9/09(W)  Lab 1.3 More plotting and (dot) m-files
9/10(Th)  *PS1.1 due. Lab 1.4 M-file Management; Differences in Cumulative Histograms
9/16(W)  Lab 1.5 Did my drug really work? Statistical inference
9/21(M)  Lab 1.6: What was the impact of my drug on average?
9/29(Tu)  *PS1.2 due. Lab 1.7: Basic sample descriptors; tinkering with plots

October 2015

10/07(W)  Lab 1.8 How well did my drug do? Variance explained and discriminability
10/12(M)  Lab 2.1: Describing data with fits (models)
10/14(W)  *PS1.3 due. Lab 2.2: Using fitting as a tool, not like a fool
10/19(M)  Lab 2.3: Creating robust index values
10/21(W)  Lab 2.4: Dealing with a truckload of data
10/26(M)  *PS2.1 due. Lab 3.1: Time series: correlation, feature detection, rates
10/28(W)  Lab 3.2: Frequency and Fourier analysis

November 2015

11/02(M)  Lab 3.3: Convolution and filtering
11/04(W)  *Team project 1 presentations
11/09(M)  *TP1 write-up due. Lab 4.1 Basic image representations
11/11(W)  *PS3.1 due. Lab 4.2: Digital image representations of the physical world
11/16(M)  Lab 4.3: Image analysis: regions of interest
11/18(W)  4.4 Image analysis: functional imaging
11/23(M)  *Team project 2 presentations
11/30(M)  *TP2 write-up due. 5.1 High-dimensional space: data points as vectors

December 2015

12/02(W)  *PS4.1, 5.2 High-dimensional space: clustering and data dimensionality reduction
12/07(M)  5.3 High-dimensional space: interpolation; how principle component analysis works
12/09(W)  Team project 3 presentations
12/14(M)  **ALL MATERIAL DUE 5pm (Team 3 write-ups, any remaining labs to be turned in)**