Data Science Tools for Vision Science Applications

with Course Instructor, Bonnie Cooper

Course Description

Data Science Tools for Vision Science Applications will provide students with an overview of current data science tools and methodology while using datasets from various fields of vision science.

Data Science is a fast growing inter-disciplinary collection of algorithms, tools and technology used to gain insights from the increasingly large and complex data generated in this digital age. Vision science and visual neuroscience have also seen a rapid increase in the volume of data generated. (e.g. neural recording arrays, imaging data, etc.) This creates a new challenge for today's graduate student: learning the field is complicated by increasingly complex data. The aim of this course is to introduce students to some of the tools and techniques from data science to meet the growing demands of today's vision science.

Why Python?

Unlike Matlab, Python is an open-source programming language known for it's readability and extensive community support. The introduction of numeric and scientific libraries such as the `scipy` stack (e.g. `numpy`, `pandas`, `scikit-learn`, etc.) have led to an increased application of python in neuroscience and psychophysics. Additionally, Python is a generalized programming language and can be utilized for diverse research applications (from experimental control to data analysis/visualization); this is an advantage over languages more specific for statistical analysis (e.g. R). Python is used extensively in industry. Therefore, exposure to the language now will be beneficial for students who may be interested in career opportunities outside of academia.

Course Objectives

1. Understand native python data structures and basic programming techniques
2. Gain familiarity to data science-centric Python libraries such as `pandas`, `numpy` and `scikit-learn`
3. Introduction to fundamentals of machine learning with Python while working with curated data sets relevant to vision science
4. Become proficient at using Git anf Github for version control
5. Build an analysis of a dataset (of the students choice) using data science tools in the Python environment.
6. Practice communicating research findings as both a written report and an oral presentation.

Course Weekly Outline

Each meeting be 2 hours long consisting of lecture and interactive coding. This schedule is subject to change depending on the progress of the class

Week	Week of	Topic	Description
1	August 17th	Python	Course Introduction & Introduction to Python (basic functionality and core data types)
2	August 24th	Python	Python - Beyond beginner Flow control and functions
3	August 31st	Python / Git	Python Classes & introductory OOP Git / Github for version control
4	Sept 7th	Data Access & Manipulation	Loading and accessing various data formats into the Python environment
5	Sept 14th	Data Access & Manipulation	the `numpy` and `pandas` libraries
6	Sept 21st	Exploratory Data Analysis & Visualization	data inspection, missingness, data cleaning
7	Sept 28th	Exploratory Data Analysis & Visualization	summary statistics and basic visualizations using `Matplotlib Pyplot` and `Seaborn`
8	Oct 5th	Experimental Design	an introduction to `psychopy` for designing and implementing behavioral experiments
9	Oct 12th	Regression	Linear, multiple, logistic regression
10	Oct 19th	Machine Learning: Supervised	basic supervised learning approaches for numeric data: regression & decision trees
11	Oct 26th	Machine Learning: Supervised	basic supervised learning approaches for categorical data: classification/categorization (e.g. kNN)
12	Nov 2nd	Machine Learning: Unsupervised	basic unsupervised learning approaches for numeric data: dimensionality reduction (e.g PCA)
13	Nov 9th	Machine Learning: Unsupervised	basic unsupervised learning approaches for categorical data: clustering (e.g. K-mean)
14	Nov 16th	Machine Learning: Neural Nets	Introduction to artificial Neural Nets with `Keras`
15	Nov 30th	Data Narratives	students will prepare a report and give a brief talk to present their findings to the class

Deliverables

Deliverable	Description	Final Grade %
Project Report	perform an analysis of a dataset using python and organize a report to summarize the results.	30%
Project Presentation	give a brief 12-15min talk to present the project findings to the class.	10%
Python Problem Set	Most course topics (e.g. Python, Regression etc) will have an accompanying problem set in the form of either a Jupyter notebook or Python script. Students will be responsible for uploading 4 completed assignments for full credit (10pnts per homework).	40%
Weekly Discussion	Each week will have a discussion topic posed to the class on Slack. Students are responsible for replying to 10 weekly discussions for full credit (1pt per post)	10%
DataCamp Courses	Several required courses will be assigned on the DataCamp platform. These courses will need to be completed by the close of the semester	10%

Course Instructor

Dr. Cooper is a post-doctoral researcher at SUNY College of Optometry in the lab of Prof. Robert M McPeek. Bonnie has experienced the graduate program at SUNYOpt having received her PhD from the program under Prof. Barry B Lee; this gives her first-hand understanding of the needs faced by the program's PhD candidates. Additionally, Bonnie is currently pursuing a Master's of Science in Data Science at CUNY with a focus on machine learning making her uniquely qualified to instruct data science methods to vision science students.