Data-Driven Research

Creative Problem Solving with Python

Me!

Hannah Wolfe
wolfe@umail.ucsb.edu
office: Phelps 3314

I have a Master's Degree in Computer Science
    and a Master's Degree in Media Arts and Technology from UCSB.

I am a PhD Student in Media Arts and Technology at UCSB.

I spent a couple of years as a Software Engineer at a solar company.

I have a Bachelor's Degree from Bennington College, concentration in Art.

My Research

Robotics
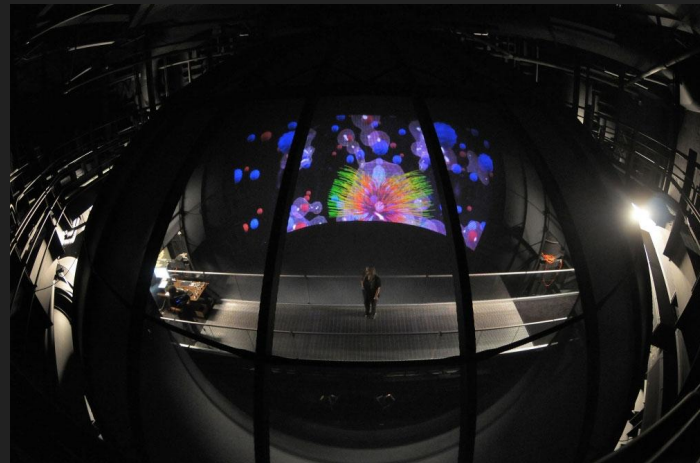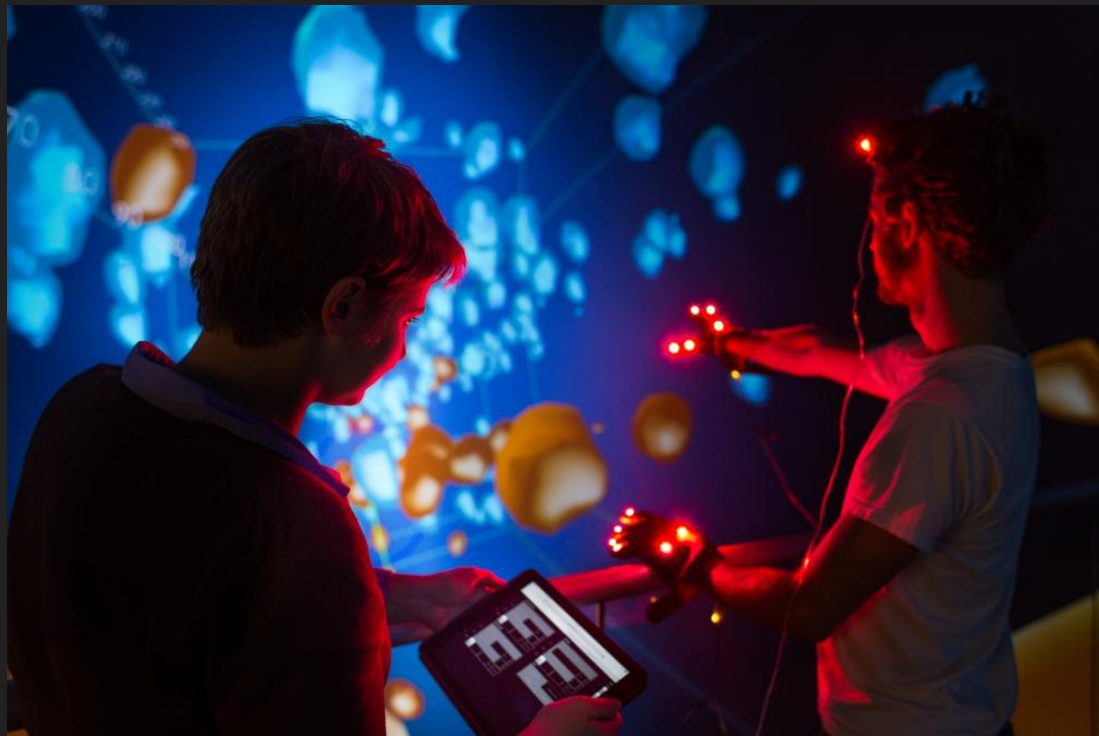
Human Computer Interaction

Affective Computing

Machine Learning

Data Visualization

Virtual Reality

AlloSphere Research Group

Syllabus

Goals of the Class

Give students skills to be able to analyze data and draw meaningful conclusions using Python.

Students will be able to...
- write a Python script that imports a dataset and uses a library.
- develop a research question and design a plan to answer it using data analytics.
- evaluate research papers critically and conduct a literature review.
- express their research in a clear and thoughtful manner verbally and in writing.
- recognize the limitations and potential pitfalls of employing machine learning techniques.

Class Format

54% Labs
  9% Regular Expression Lab
  9% Data Formats Lab
  9% Data Visualization Lab
  9% Unsupervised Learning Lab
  9% Supervised Learning Lab
  9% Neural Networks Lab

35% Research Project
  5% -Research Assignment 1
  5% -Research Assignment 2
  5% -Research Assignment 3
  10% -Final Presentation
  10% -Final Report

11% Participation

Grading Policy

- All assignments will be submitted through Gauchospace due at 9 am.
- Students are allowed 2 extension days for <u>individual work</u>
    on one assignment <u>OR</u> 1 extension day on two assignments.
- To use a late day please email both the TA and Professor before the deadline.
- Late work will not be accepted without Professor/TA's approval.

Course Policies

- Attending Lecture, Labs and Sections are mandatory.
- Any form of harassment is unacceptable.
- Please bring your cell phone to class charged and on silent.
  Cell phones may only be used during lecture when the instructor allows it.
- Computer use is only allowed with instructor consent.
- Reliable email/slack communication is required, emails with questions must be responded to in a timely fashion.
- Read and adhere to the requirements for academic integrity.
- Bring a notebook and pencil to class daily.

Scientific Method?

Discovery-Driven VS. Hypothesis-Driven Research

Hypothesis-driven data mining

"Begins with a proposition by the user, who then seeks to validate the truthfulness of the proposition"

Discovery-driven data mining

"Finds patterns, associations, and relationships among the data in order to uncover facts that were previously unknown or not even contemplated by an organization"

Rahman, Hakikur, ed. Social and Political Implications of Data Mining: Knowledge Management in E-Government: Knowledge Management in E-Government. IGI Global, 2009.

Poor Science: Data-driven Science Done Wrong

No hypothesis.

Assumptions: alternative models not explicitly examined.

Statistical basis of model either neglected or only implicit.

No cross-validation: just one form of evidence.

Using greedy algorithms, sensitive to noise.

Measures of significance weak or absent, both computationally and experimentally.

Data-driven Science Done Right

Multiple competing hypotheses.

Alternative models explicitly included, computed, to eliminate assumptions.

Statistical models clear, well-justified.

Multiple, independent types of evidence.

Robust algorithms w/ well demonstrated convergence to global optimum.

Rigorous posterior probability calculated for all possible models of the data.

Machine Learning

Machine learning gives "computers the ability to learn without being explicitly programmed."

-Arthur Samuel in 1959

Data-Driven Research Applications?

Data-Driven Research Applications

Biology/Biotechnology/Healthcare

Astronomy/Earth Science

Politics

Human Geography

Urban Planning

Social Networks

Sports

Financial Analysis

Robotics

Natural Language Processing

Sentiment Analysis

Psychology

Project

Through this project we will...

1. Choose a subject
2. Literature Review on subject
3. Define a question
4. Choose a dataset
5. Analyze the dataset
6. Synthesize results
7. Draw conclusions from results

Project

Research Assignment 0: Choose General Topic

Research Assignment 1: Literature Review

Research Assignment 2: Choose Dataset and Define Question

Research Assignment 3: Data Analysis

Research Presentation

Research Paper

Research Assignment 0

Due tomorrow by 9am on Gauchospace

- Choose your group's general topic
- Set-up Gauchospace and write a sentence about your chosen topic
- Set-up Slack and upload a picture

Project Groups

Group 1:
Chatura
Chinmayi
Brandon
Nicholas Ng

Group 2:
John
Ian
Sahana
Nick Lewis

Group 3:
Jack
Neil
Arleen
Rohit

Group 4:
Raymond
Oscar
Riley
Sohel

Group 5:
Adrian
Gordon
Cassidy
Marika

Group 6:
Beck
Sagnik
Kai
Darren