

Syllabus

Course Number and Title: CS 640-01 Bioinformatics 4 units

Location: LM 346A

Schedule: 4:45 pm-06:25 pm

Instructor: [Patricia Francis-Lyon](#)

Email: pfrancislyon@cs.usfca.edu

Office: Cowell 203

Office Hours:

In-person: Mon 3-4 pm in CO 203,
Mon, Wed 8:15 - 9:15 pm in LM 346 B

Virtual (via Zoom) <https://zoom.us/j/263772282>
Tues, Thurs 6-7 pm,
Sat: 12 noon - 1 pm

TA (for grading): Chiral Mehta cmehta2@dons.usfca.edu

Receiving email

I set aside time nearly every day to assist students during office hours (online or in-person). Generally, you should ask your questions in class or at these times rather than via email.

Email messages of a more personal nature should be sent from your UNIVERSITY OF SAN FRANCISCO email address. Expect replies to be delayed on weekends, holidays and during conference attendance.

Course Description:

This course will introduce you to the field of bioinformatics. Assuming proficiency in computer programming but little or no exposure to Molecular Biology and genomics, this course will bring you up to speed so you can do research involving proteins, DNA, and RNA. The principles underlying analysis of genomic and proteomic data will be delineated. Students will apply bioinformatics knowledge and algorithms in writing efficient scripts to analyze data from cancer research, high throughput sequencing and other domains. You will gain practical experience in the use of bioinformatics tools and algorithms and will utilize APIs for scientific programming, data analysis and visualization.

Approach

This is a lab-based and projects-based class. Students will become proficient in the use of bioinformatics tools and the writing of scripts through hands-on labs and assignments. Students will choose a project of their interest for in-depth skill development. Projects range from deep learning for predictive analytics with EHR data to analysis and visualization of cancer data using R.

Course Objectives: *Upon completion of the course, the student will:*

1. Explain the role of DNA and proteins in human disease and therapies.
2. Analyze biological data using tools such as Weka, the R programming language, and Python/Pandas.
3. Appraise, select and utilize relevant data from databases such as Genbank, the Protein Data Bank, the Cancer Atlas and the Gene Expression Omnibus to solve a given problem.
4. Apply bioinformatics tools such as BLAST and multiple sequence alignment to work with data encoding genes, proteins, DNA, and RNA.
5. Solve problems independently, utilizing documentation and code from APIs such as Numpy, Pandas, Scikit-learn, and the Bioconductor suite of tools in R.
6. Utilize machine learning algorithms such as neural networks and support vector machines for scientific programming and data analytics.

List of Topics: (for every topic, focus is on computation and algorithms)

- Genomics:
 - DNA -> RNA -> proteins, open reading frames in Python
 - data: Genbank: prokaryote, eukaryote
 - tools: BLAST, MSA, R
 - genetic diseases
- Proteins: sequence, structure and function. Representation and visualization
 - data: Protein Data Bank (RCSB): .pdb files
 - tools: Chimera
 - protein folding and related diseases
- Profiles / PsiBlast / Hidden Markov Models & Viterbi algorithm in Python
- Data analysis and Visualization with Python/Pandas, R
- Analysis, Classification, Machine Learning, Metrics with Python/Pandas, R, Weka
- High throughput Sequencing: NGS, RNA-Seq in R
- Personalized Medicine: human genome and variants
 - data: 23andme, dbSNP: .vcf files
 - tools: PLINK
- Human Disease

Suggested Bioinformatics Textbook

Bioinformatics and Functional Genomics 3rd edition, Jonathan Pevsner, Oct 2015

Recommended Readings

Genomics and Personalized Medicine: What Everyone Needs to Know®, Michael Snyder, 2016

Each student will have a free account with Datacamp for 6 months. Datacamp has many resources for learning R and Python/Pandas

There are many free online resources for learning Pandas. Here are two:

Pandas Tutorials from pandas.pydata.org: <http://pandas.pydata.org/pandas-docs/stable/tutorials.html>

Python Pandas Tutorial from tutorialspoint.com:

https://www.tutorialspoint.com/python_pandas/python_pandas_introduction_to_data_structures.htm

Here are some texts (not required) for learning R:

Peng, Roger, (2016). **R Programming for Data** <https://leanpub.com/rprogramming>

Davies, Tilman, (2011). **The Book of R: A First Course in Programming and Statistics**

Assignment Description and Grading Criteria

<i>Assignments</i>	<i>% of semester grade</i>
Python/R labs	5%
Bioinformatics Assignments	35%
Quizzes /quick reports (lowest 1/3 are dropped)	10%
Midterm	20%
Presentation & Participation	5%
Project	25%
Total	100%

Grading Scale

<i>Percentage (%) Score</i>	100	99-92	91-90	89-88	87-82	81-80	79-78	77-75	74-70	69-0
<i>Grade</i>	A+	A	A-	B+	B	B-	C+	C	C-	F

Assignments and Assessments:

A quiz or a quick report will occur about once per week.

An assignment or lab will be due almost every week.

You will undertake a semester-long bioinformatics project and present it in class.

You are expected to participate in discussions, including presentations by fellow students.

An in-class midterm will be administered (see below).

Midterm

All students must be present for the midterm exam on the date it is scheduled. **If you cannot be present for an exam, you must notify the faculty, well in advance of the scheduled exam, and arrange to take a comparable exam before the scheduled exam.** Make-up exams are available only to students who experience documented (ex: doctor's note) emergency situations within the academic semester. If you are eligible for a make-up exam, it must be scheduled prior to the next class meeting.

Academic Integrity:

In this course you must do your own work on exams and assignments, unless explicitly specified otherwise. Any substantial contribution from another person or an online source must be cited to avoid plagiarism. Code-walkthroughs are an accepted practice in computer science, and every student can expect to engage in review of their own code. Doctoring test cases to make it look as if your program works correctly is cheating. Test cases must be actual copy-pastes of the results of your run.

Academic Integrity goes beyond cheating and plagiarism. In order to foster an environment of fairness, the University commits to the standards of academic behavior listed in the Honor Code found at the [Honor Code](http://myusf.usfca.edu/academic-integrity/honor-code) page (<http://myusf.usfca.edu/academic-integrity/honor-code>) under the [Academic Integrity](http://myusf.usfca.edu/academic-integrity) website (<http://myusf.usfca.edu/academic-integrity>).

Course Grade Appeal

Information about the grade appeal process can be found at the [Appeals Process for Change of Course Grade](https://www.usfca.edu/catalog/policies/appeal-process-change-of-course-grade) page (<https://www.usfca.edu/catalog/policies/appeal-process-change-of-course-grade>).

Students with Disabilities:

If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF [Student Disability Services \(SDS\)](https://www.usfca.edu/student-disability-services) (<https://www.usfca.edu/student-disability-services>) at 415 422-2613 within the first week of class, or immediately upon onset of disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit www.usfca.edu/sds or call (415) 422-2613.