SYLLABUS

Course title and number	Single-cell Data Analysis via Machine Learning, VIBS-689
Term	Spring 2020
Meeting times and location	Tue & Thr 10 AM - 12 PM, VIDI (bldg #1813) rooms 104 & 115

Course Description and Prerequisites

Principles and concepts in single-cell RNA sequencing (scRNAseq) experiments; real-world applications of scRNAseq with examples; machine learning (ML) methods for single-cell data analysis; practical and effective ML methods and concepts; applications of ML methods in high-dimensional scRNAseq data; algorithm design and development of scientific software using high-level high-performance scientific computer languages; emerging techniques for integrative single-cell data analysis, and the assumptions, advantages, and limitations of these techniques.

Prerequisites: Instructor approval.

This course is designed for multi-disciplinary education. Enrollments are encouraged from students with backgrounds in any of the following disciplines: biology, genetics, biochemistry, medicine, and ecology. Students with backgrounds in computer science, statistics, and mathematics who want to learn how to apply ML methods to scRNAseq data are also encouraged.

Learning Outcomes or Course Objectives

By the end of the course, students are expected to understand:
- Practical analytical framework for scRNAseq data
- A core set of practical and effective ML methods and concepts
- Principle of effective experimental design and analysis of the results
- The anatomy of analytical workflow and procedure of workflow design

Students should be able to:
- conduct scRNAseq data analysis with real data
- extract information from multiple scRNAseq data sets for comparative analysis
- apply existing ML algorithms to their own experimental data to solve real-world problems
- design new ML algorithms conceptually based on existing ones
- explore the common pitfalls in interpreting statistical arguments, especially those associated with scRNAseq data

Instructor Information

Name	James Cai
Telephone number	979-458-5482
Email address	jcai@tamu.edu
Office hours	Wed 10 AM -12 PM

Office location   VMR (bldg. #1811) room 384

Course Topics, Calendar of Activities

Week	Topic	ML lab
1	Introduction to single-cell technologies (10x Genomics and C1)	Mathematical prerequisites and programming languages
2	Getting expression counts (UMIs) for individual cells (10X cell-ranger and salmon-alevin pipeline)	Vectors and matrices: linear algebra
3	Basic statistics of scRNAseq data and QC	Probability distributions: statistics review
4	Modeling and data normalization	Distances, anomaly detection, and curve fitting
5	Dimension reduction and visualization	Dimension reduction: PCA, tSNE, UMAP, PHATE
6	Clustering analysis of high-dimensional scRNAseq data	Clustering: k-means, DBSCAN, spectral clustering
7	Marker genes and cell type identification	Classification: classifiers, LDA, random forest, k-NN, neural network
8	Feature selection, identification of highly variable genes	Kernel methods: support vector machine
9	Differential expression analyses with scRNA-seq data	Optimization: gradient descent, evolutionary algorithm
10	scRNAseq vs bulk RNAseq	Components and decomposition: SVD, ICA, NMF
11	Pseudotime and trajectory analyses	Correlation and regression: multivariate analysis, regularization, LASSO
12	Manifold alignment for combining different types of data, e.g., scRNAseq and scATACseq	Tensor factorization and manifold learning
13	Construction of single-cell gene regulatory networks (scGRNs)	Network and community: centrality analysis, community detection, clustering with multi-layer graphs
14	Student presentations	

Attendance and Grading Policies

The University views class attendance as the responsibility of an individual student. Attendance is essential to complete the course successfully. University rules related to excused and unexcused absences are located online at http://student-rules.tamu.edu/rule07

1. Homework Assignments (50%) [homework score sheet]
        Consists of programming tasks or short written assignments handed out every
        week.
2. Presentation (25%)
        Towards the end of the semester, each student will give a 15-min platform
        presentation.
3. Final Exam (25%), 30 multiple choice questions to be completed in 50 mins.

Late work will be accepted only if there is an approved absence, http://student-rules.tamu.edu

Class Notes and Resource Materials

Class notes are distributed in class. There is no required textbook.

Required software – R, Matlab and Julia:
1. R and Julia are free software.
2. Matlab is available on computers in VIDI 115 (via the VOAL) and on computers in the Open Access Labs across the campus.

Americans with Disabilities Act (ADA) Policy Statement

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact Disability Resources in the Student Services Building or at (979) 845-1637 or visit http://disability.tamu.edu. Disabilities may include but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability-related needs with Disability Resources and their instructors as soon as possible.

Academic Integrity

For additional information, please visit: https://aggiehonor.tamu.edu

"An Aggie does not lie, cheat, or steal, or tolerate those who do."

Upon accepting admission to Texas A&M University, a student immediately assumes a commitment to uphold the Honor Code, to accept responsibility for learning, and to follow the philosophy and rules of the Honor System. Students will be required to state their commitment to examinations, research papers, and other academic work. Ignorance of the rules does not exclude any member of the TAMU community from the requirements or the processes of the Honor System.