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
Course Topics, Calendar of Activities

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

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](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](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](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](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.