Georgia Institute of Technology / Department of Biomedical Engineering

BMED6517 Machine Learning in Biosciences

Instructors: Peng Qiu
Credit: 3-0-3
Prerequisites: (MATH 2401 or MATH 2411 or MATH 2605) **AND** (CS 1332 or CS 1371)

Text Books:
Trevor Hastie, Robert Tibshirani, Jerome Friedman. “The Elements of Statistical Learning”, 2nd Edition.

Description
This course introduces machine learning concepts and methods. The course is targeted to graduate students in Biomedical Engineering, Bioinformatics, Computer Science, Electrical Engineering and related disciplines, and also qualified undergraduate students. Topics to be covered include supervised and unsupervised learning, dimension reduction and visualization. Topics will be accompanied by research papers in the bioinformatics domain, relating algorithms to biological applications.

Targeted Students
The course is designed to be at an introductory level, and target on students who have basic math and linear algebra, but have not taken Stochastic Process or other advanced Statistics courses. This course can benefit students in BME, BIOE (ECE or CS), and bioinformatics.

Objectives and Expected Outcomes
The course aims to provide an introduction to the basic principles and techniques of machine learning, and its applications in biological data analysis. Projects are intended to relate computational algorithms to biological applications, involving program implementations of algorithms in Matlab (or R) and analysis of real datasets. By the end of the course the students should:
1. Understand the basic concepts in machine learning.
2. Have mastered standard techniques for supervised and unsupervised learning algorithms.
3. Understand the formulation and challenges of machine learning problems in biological applications.
4. Have acquired skills for implementing a set of machine learning algorithms for their own applications.

Grading:

Early-term project	15%
- Predictions due:	end of week 06 (15%)
- Presentation:	week 07 (0%)
Term Project	85%
- One page proposal:	week 06 (25%)
- Project update:	week 12 (25%)
- Report/presentation:	week 16 (35%)
No Final exam	

Minimum grade cutoffs are:

>80%	A
>60–80%	B
>50–60%	C
>40–50%	D
<40%	F

Early-term project: Designed by instructor.
Term project: Students are encouraged to propose term projects relevant to their own research.
Proposal: 1-page, motivation, problem description, data, method, expected outcome, timeline
Project update: 1-paragraph description of progress (schedule a meeting if discussion is needed)
Final report: 4-pages, conference paper format (Abstract, Intro, Method, Results, Conclusion)
NOTE: Projects are done in teams of 1~3 students. Students on the same team receive the same grade.

Tentative Topical Outline

	Date	Topic	Due
Week 01	1/18	Martin Luther King, Jr. National Holiday	
	1/20	Introduction of machine learning and applications Examples of previous semesters' projects	Start thinking about final project
Week 02	1/25	Introduction to Early-term Project (AML Classification using flow cytometry)	
	1/27	Review of probability, linear algebra. Data Resources Examples of previous semesters' projects	
Week 03	2/1	Supervised learning framework, KNN, linear classifier	
	2/3	Linear regression and regularization, Facilitate team formation	
Week 04	2/8	Linear and Logistic Regression for Classification	
	2/10	Feature selection (ttest, var, pvalue, FDR) Hints on Early-term project	
Week 05	2/15	Support Vector Machine	
	2/17	Decision Tree and Random forest	
Week 06	2/22	Model assessment (CV, P-R, ROC, AUC, F1, overfit)	Final project one-page proposal due
	2/24	Neural Networks, Deep Learning	Early-term project prediction due
Week 07	3/1	Early-term Project Presentations	
	3/3	Early-term Project Presentations	
Week 08	3/8	Clustering algorithms (hierarchical, kmeans, ensemble clustering)	
	3/10	Clustering algorithms (mixture models, GMM, spectral clustering)	
Week 09	3/15	Clustering algorithms (community finding, DBSCAN, density peaks)	
	3/17	Dimension Reduction and Visualization (PCA, isomap, spring embedding)	
Week 10	3/22	Dimension Reduction and Visualization (tSNE, UMAP, TPE, SPADE)	
	3/24	Mid-semester break from instruction	
Week 11	3/29	Advanced topics: Single-cell sequencing for cell subtype identification (Seurat)	Final Project update
	3/31	Advanced topics: Single-cell sequencing for finding trajectories associated to biological progression (TPE)	
Week 12	4/5	Advanced topics: Imputation of single-cell RNA-seq data (MAGIC and scImpute)	
	4/7	Advanced topics: Embracing the dropouts in single-cell RNA-seq data (co-occurrence clustering)	
Week 13	4/12	Advanced topics: Cell type mapping	
	4/14	Advanced topics: Robust correlation	
Week 14	4/19	Final Project Presentations	
	4/21	Final Project Presentations	
Week 15	4/26	Final Project Presentations	Final project report 4-pages