Machine Learning for Biomedical Applications

Department of Biomedical Engineering
University of Alberta, Edmonton
BME 677
Class periods: To be decided
Location: To be announced
Academic Term: Fall 2023

Instructor:

Maral Aminpour, Ph.D. (Pronouns: she/her/hers)
Email address: aminpour@ualberta.ca
Office hours: To be decided
Office location: Donadeo Innovation Centre for Engineering 13-209

Course Description:

Applications of machine learning tools to real-world problems in biomedical engineering including diagnostic and prognostic applications. An introduction to machine learning. Machine learning tools: regression and classification; manifold learning and dimensional reduction; decision trees and ensemble learning; unsupervised learning and clustering; feature selection and feature extraction; neural networks and deep learning. Biomedical applications: cancer, cardiovascular disease, diabetes, neurological diseases and infectious diseases.

Course Objectives and General Content:

This course is primarily intended for graduate students in the Department of Biomedical Engineering (BME) and it will also be available to qualified graduate students in other engineering disciplines and students from other faculties (for example, Science and Medicine & Dentistry), and interested undergraduate students with the permission of the instructor.

The course focuses on applications of recent machine learning technologies that deliver promising solutions and improved technologies in the biomedical field. The complexity of the analysis and the enormous repository of data in healthcare means that machine learning will increasingly be applied within the field. Machine learning algorithms can learn patterns from biomedical data sets in ways that are different and sometimes better from human cognition and provide insights on disease diagnosis and prognosis by decoding previously hidden correlations among symptoms, vital signs, and disease. In this context, the goal of the proposed course is to introduce the application of machine learning tools to real-world problems in biomedical engineering. The emphasis of this course is on practical applications of machine learning in biomedical engineering, but with the recognition that it is impossible to be an intelligent practitioner without some appreciation of underlying theory. The course will cover an overview of a wide range of machine learning tools (such as regression, classification, decision trees, random forests, clustering, artificial neural networks and convolutional neural networks), biomedical data sets (image data, omics data and patient records) and disease applications (including cancer, cardiovascular disease, diabetes, neurological diseases and infectious diseases). There will be hands-on tutorials and a final group project.

This course provides self-contained material that will take a beginner in Machine learning from learning the basic Python commands, through understanding of classical machine learning techniques and implementing them in Scikit-learn, all the way towards the state-of-the-art deep learning in Py-Torch. However, the knowledge of basic programming principles in any programming language, as well as the basic linear algebra, calculus and probability theory, will allow the reader to fully absorb the materials in this course and gain in-depth understanding and confidence. A Python boot camp will be provided in the first week of the course covering the key elements of Python programming. Training will be provided during the course on the Python libraries needed. For readers with no programming background, we recommend complementing this course with one of the popular online courses or books on Python for beginners.

Learning Outcomes:

By the end of the course the student will be able to:

1. Understand machine learning techniques, including regression and classification; manifold learning and dimensional reduction; decision trees and ensemble learning; unsupervised learning and clustering; feature selection and extraction; artificial neural networks and deep learning.
2. Understand both the advantages and limitations of each technique with respect to biomedical problems and know how to select appropriate techniques for specific biomedical problems, including diagnostic and prognostic tasks.
3. Be able to train models in Python, Scikit-learn and PyTorch API; perform model selection, evaluation and troubleshooting; apply best practice implementation and generalisation to new data
4. Critically analyse related recent literature on machine learning for BME applications; write reports and give oral presentations on machine learning topics.

Marking Scheme:

Assignment	**Topics**	**Deadline**	**Percentage of Final Grade**
Assignment 1	Python and machine learning	Monday September 25 8:00 am	12%
Assignment 2	Regression and classification	Monday October 9 8:00 am	12%
Assignment 3	Manifold learning and dimensional reduction; Decision trees and ensemble learning	Monday October 23 8:00 am	12%
Assignment 4	Unsupervised learning, clustering, feature selection and feature extraction	Monday November 6 8:00 am	12%
Assignment 5	Neural networks and deep learning	Monday November 27 8:00 am	12%
Group project	*Selected by students*	Presentation: December 1–7	40%

		Report and source code: Thursday December 7 6:00 pm	

There will be five individual assignments and one group project.

Each individual assignment will correspond to topics covered in class lectures. There will be a total of five assignments with the deadlines listed in the course schedule. Each assignment will be submitted as an iPython notebook file (.ipynb). Detailed marking rubrics for the assignments will be provided on GitHub and eClass.

There will be a final group project due at the end of the course. A list of suggested topics is included in the syllabus. The groups and project topic must be approved by the instructor. Students must submit a project proposal to the instructor by the deadline in the course schedule. The project will require reading of relevant literature and the creation of Python code. The project will require three components to be submitted: (1) a presentation during the last week of classes; (2) a written report detailing the problem, the methods used, and a discussion of the results; and (3) an archive containing all Python source code and data files. Detailed requirements and marking rubrics for the final project will be provided on the course GitHub and eClass.

There will be no final exam.

Textbook and Course Materials:

Recommended textbook:
Machine Learning for Biomedical Applications: With Scikit-Learn and PyTorch by Maria Deprez Emma C. Robinson; September 15, 2023. ISBN-10: 0128229047. ISBN-13: 978-0128229040.

Publisher description: "*Machine Learning for Biomedical Applications* presents machine learning techniques most commonly used in a biomedical setting. Avoiding a theoretical perspective, it provides a practical and interactive way of learning, where concepts are presented in short descriptions followed by solving simple examples using biomedical data. Interactive Python notebooks are provided with each chapter to complement the text and aid understanding."

Required software:
The course will require Python 3. Interactive Python (IPython) notebooks will be used. The following Python libraries will be used: NumPy, SciPy, Scikit-Learn, Matplotlib, Pandas and PyTorch. All software to be used is open source and freely available.

Other course materials:
A GitHub repository will contain all lecture presentation slides, tutorials, IPython notebooks, data files, and all other course files. The files will also be available on eClass.

Course Content:

Note: The course schedule might be subject to change.

Week (dates)	**Subject**	**Biomedical Applications**	**Deadlines**
1 (Sep 1–7)	Python programming boot camp		
2 (Sep 8–14)	Introduction to machine learning		
3 (Sep 15–21)	Regression	Predict neonatal brain growth	Group project topic approval Thu Sep 21
4 (Sep 22–28)	Classification	Diagnosis of heart failure	Assignment 1 due Mon Sep 25
5 (Sep 29–Oct 5)	Manifold learning and dimensional reduction	Analysis of functional magnetic resonance imaging (fMRI)	
6 (Oct 6–12)	Decision trees and ensemble learning	Predicting diabetic retinopathy; Identifying synergistic drug combinations	Assignment 2 due Mon Oct 9
7 (Oct 13–19)	Unsupervised learning and clustering	Breast cancer diagnosis; Classifying cancer subtypes	
8 (Oct 20–26)	Feature selection and extraction	Prediction of age at scan from volumes of brain structures of preterm babies.	Assignment 3 due Mon Oct 23
9 (Oct 27–Nov 2)	Artificial neural networks (ANNs)	Classification of tumors in chest/lung CT scans and differentiate between benign and malignant tumors	
10 (Nov 3–9)	Deep learning and convolutional neural networks (CNNs)	Detecting COVID-19 pneumonia from X-ray images	Assignment 4 due Mon Nov 6
11 (Nov 10–23)	*Additional topic selected by instructor*		
12 (Nov 24–30)	*Additional topic selected by instructor*		Assignment 5 due Mon Nov 27
13 (Dec 1–7)	Student project presentations		Group project due Thu Dec 7

The outline of the course the lectures is adopted from Dr. Maria Deprez and Dr. Emma C. Robinson based on *"Machine Learning for Biomedical Applications: With Scikit-Learn and PyTorch" book.*

Week 1: Python programming boot camp
A brief introduction to Python 3 for beginners, including Data types. Operations. Control flow. File I/O. Python packages and Modules : NumPy for linear algebra, Matplotlib for plotting and Pandas for data handling. An introduction Python Notebooks and provide instructions for installation of Anaconda software package.

Week 2: Introduction to machine learning
Basic machine learning concepts. Using the Scikit-Learn library. Training machine learning models: performance measures; overfitting and underfitting; training, testing and cross-validation.

Week 3: Regression
Multivariate linear regression; normal and gradient descent. Penalized linear regression; ridge and lasso regression; regularization. Non-linear regression; feature transformation; kernel trick. ***Biomedical Application:*** prediction of neonatal brain growth (we demonstrate the problem of overfitting on example of prediction of age of a baby from brain volumes extracted from MRI scans and introduce of the kernel trick.)

Week 4: Classification
Linear classification; binary classification; multi-label classification. Support vector classification; linear support vector classification; non-linear support vector classification; kernel support vector classification.
Biomedical Application: diagnosis of heart failure. We demonstrate these concepts on an example of prediction of heart failure using the features extracted from cardiac MRI, ultrasound and ECG.

Week 5: Manifold learning and dimensional reduction
Principal component analysis (PCA). Independent component analysis (ICA). Non-linear manifold learning and Laplacian eigenmaps.
Biomedical Application: analysis of fMRI data. We show how reducing abdominal MRI to two dimensions allows us to extract the breathing cycle for correction of the motion artefacts.

Week 6: Decision trees and ensemble learning
Weak learners and decision stumps. Entropy, information gain and Gini index. Decision trees. Ensemble learning: voting, bagging, random forests and boosting.
Biomedical Application: predicting diabetic retinopathy using Decision trees and identifying synergistic drug combinations using RF.

Week 7: Unsupervised learning and clustering
Unsupervised learning. Clustering algorithms: K-means clustering and gaussian mixture model. Spectral clustering using Laplacian eigenmaps.
Biomedical Application: Diagnosing breast cancer from biopsy images

Week 8: Feature selection and extraction

How and why to select features. Feature selection methods: univariate, model-based and recursive feature elimination. Feature extraction.
Biomedical Application*:* Prediction of age at scan from volumes of brain structures of preterm babies (selection from 5 features)

Week 9: Artificial neural networks (ANNs)
Introduction to basic building blocks of neural networks, including the artificial neuron and single-layer perceptron. We cover activation functions, loss functions for regression and classification, and training of single layer neural network models. We cover training of neural networks using backpropagation. We provide a Pytorch tutorial and implement neural network regression and classification examples in Pytorch
Biomedical Application: Classification of tumors in chest/lung CT scans and differentiate between benign and malignant tumors using NN & Transfer Learning

Week 10: Deep learning and convolutional neural networks (CNNs)
Introduction to fully Connected Neural Networks: how multiple linear layers and non-linear activation functions are combined to create deep neural network architectures. We will describe a more advanced deep learning architecture, the convolutional neural network.

B*iomedical Application (deep learning):* We show a complete Pytorch deep learning solution for a real world biomedical problem for detecting COVID-19 pneumonia from X-ray images.
And/or
B*iomedical Application (CNN):* we will present an example of segmentation of neonatal brain MRI into multiple anatomical regions using U-Net architecture, while demonstrating how to handle large datasets in Py-Torch.)

Weeks 11–12: Additional topics selected by instructor
To be decided.

Week 13: Student project presentations
Time allocated for student project presentations.

Example projects:

1. Predict survival outcome in breast cancer patients using clustering
2. Classification goal is to predict whether the patient has 10-year risk of future coronary heart disease using logistic regression
3. Breast Cancer biopsy image analysis using a Convolutional Neural Network (CNN)
4. Diagnosis of Parkinson's using voice measurements
5. Urinary Tract Infection prediction from health record data
6. Analysis of COVID-19 transmission in South Korea
7. Predicting survival outcome in breast cancer patients based on their gene expression

Students can choose other topics for their projects after discussing it with the instructor and getting the approval of the instructor.

Policies and Student Responsibilities:

Attendance Policy, Assignment Expectations, and Make-Up Policy

Attendance in classes is optional. Assignments are due on the deadlines listed in the course schedule. **No late assignments will be accepted.** Missed assignments can only be made up for in extraordinary circumstances with prior approval of the instructor.

Students Requiring Accommodations

Accommodations are available by contacting Academic Success Center (https://www.ualberta.ca/current-students/academic-success-centre/accessibility-resources/index.html) and by providing appropriate documentation. Once registered, an eligible student will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.

Academic Success Centre

The Academic Success Centre (https://www.ualberta.ca/current-students/academic-success-centre/index.html) provides professional academic support to help students strengthen their academic skills and achieve their academic goals. Individual advising, appointments, and group workshops are available year-round in the areas of Accessibility, Communication, Learning, and Writing Resources. Modest fees apply for some services.

Commitment to a Safe and Inclusive Learning Environment

The Faculty of Engineering values broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Resource Center (https://policiesonline.ualberta.ca/policiesprocedures/policies/discrimination-harassment-and-duty-to-accommodate-policy.pdf) and (https://www.ualberta.ca/human-resources-health-safety-environment/employment-information/ethical-conduct/discrimination-and-harassment.html) by providing appropriate documentation). If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following: For more information about disclosures, contact OSDHR by email at osdhr@ualberta.ca or telephone at 780-492-7357. More information about Office of Safe Disclosure and Human Rights (OSDHR) (https://www.ualberta.ca/services/provost/disclosing-discrimination-harrassment-and-other-misconduct.html).

Academic Integrity

According to the International Centre for Academic Integrity, University of Alberta's students are bound by academic integrity which states "a commitment to five fundamental values: honesty, trust, fairness, respect, and responsibility," (Fishman, 2013, p. 16). Students should obey the academic integrity on all work they submit for credit. The academic integrity source (https://www.ualberta.ca/centre-for-teaching-and-learning/teaching-support/assessment/academic-integrity.html) explains the behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns,

please consult with the instructor or TAs in this class. Check academic integrity during COVID-19 (https://www.ualberta.ca/current-students/academic-resources/academic-integrity/index.html)/

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. Resource: https://policiesonline.ualberta.ca/policiesprocedures/policies/information-technology-use-and-management-policy.pdf