Course Outline: **Machine Learning for Biomedical Data**

Course #: BMDE 520
Term: Fall
Year: 2020
Course pre-requisites: MATH 203, MATH 223, MATH 323, and one course in computer programming (or equivalent background with permission of instructor). Additional background in Calculus and Probability is recommended, e.g. MATH 222 and MATH 204, but not required.
Course co-requisite(s): None
Course schedule (day and time of class): Tuesdays, 4-7 p.m. EDT
Number of Credits: 3

Course Location: Remote (Zoom)

Course Instructor: Dr. Danilo BZDOK, Associate Professor, Department of Biomedical Engineering
Email: danilo.bzdok@mcgill.ca
Telephone number for office appointments: 514-398-6738
Office location: Duff Medical Building, Room 310
Office hours for students: TBD

Course Overview

Modern datasets are rapidly increasing in size and complexity in industry and government, and especially in the many branches of biomedicine. This course will provide a conceptual overview of practically relevant aspects of different modeling tools, mostly focusing on classical machine learning and deep learning, while comparing these to key principles of mainstream statistics and Bayesian modeling. The goal will be to provide key modeling intuitions, rather than diving deep into details of mathematical proofs and theorems, while we will contemplate outcomes and consequences of the important formal properties of the discussed modeling tools. We will focus applications and examples on the recently emerged high-dimensional biomedical measurements, centering on high-resolution body imaging and whole-genome common variant genetics, while also tapping into other modalities of the expanding richness of biomedical information. The majority of discussed tools are applicable and effective in various scenarios that involve modern datasets with many variables per observation.

Course Description

Theoretical and practical course in machine learning applied to the expanding richness of biomedical data, including multi-dimensional biomedical measurements centering on high-resolution body imaging and whole-genome common variant genetics.

Learning Outcomes
By the end of this course students are expected to:
1) Understand core properties that distinguish different classical and emerging machine learning tools and biomedical data.
2) Understand and distinguish between several practical examples of how biological research questions and specific analysis tools are applied to biomedical data.
3) Appreciate the nuances of how distinct quantitative findings and conclusions are enabled by different modeling tools, especially in relation to biomedical research.
4) Demonstrate practical know-how for building machine-learning analysis pipelines in Python.

Instructional Method
Lectures and weekly programming exercises with analysis cases in Python

Required Course Materials
Course material, prepared by the Lecturer, will be available to registered students via MyCourses.

Course Content / Outline

Week 1: Vocabulary

- Formal and informal definitions
- Application examples
- Crash course on common technical terms used in machine learning (ML)
- ML versus data-mining
- ML versus statistics
- ML vs artificial intelligence
- Wide data versus long biomedical data

Required reading:
  - Domingos, P., 2012. A Few Useful Things to Know about Machine Learning. Communications of the ACM 55, 78-87.
  - Bzdok, D., Krzywinski, M., Altman, N., 2017. Machine learning: a primer. Nature Methods 14, 1119-1120.

Week 2: Bootstrap

- The bootstrap as a case example for computer-intensive statistics
  - This example is becoming especially interesting now with richly sampled digital-sensor data, questionnaire profiling, biomedical data repositories like body imaging and GWAS approaches in genetics
- History, definition, and answering how bootstrapping extends the toolbox of statistical methods
- The different modelling purposes of the bootstrap
- Procedure
- Point estimates versus interval estimates
- Properties including non-parametric behavior
- Relation to jackknife
- Comparison to MCMC and Bayesian statistics

Required reading:
  - Efron, B., Tibshirani, R.J., 1994. An introduction to the bootstrap. CRC press. (*first chapters*)
  - Efron, B., Tibshirani, R.J., 1991. Statistical data analysis in the computer age. Science 253, 390-395.

Week 3: Curse of dimensionality (as a consequence to rich data acquisition)

- Introduction by anomaly detection
- Unit cube as a function of increasing variables p
- Sampling density as a function of N^(1/p) for population of the input space
- Examples of k-nearest neighbor
- Principal component analysis and linear regression go awry in high dimensions,
- 3 practical "solutions" to the curse in examples (dimensionality reduction, sparsity, domain-informed bias)

Required reading:

- Domingos, P., 2012. A Few Useful Things to Know about Machine Learning. Communications of the ACM 55, 78-87.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA 103, 3863-3868.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Heidelberg, Germany. (*chapter 16 / high dimensions*)

Week 4: Sparsity

- Linear regression versus Lasso
- Lasso versus Ridge versus ElasticNet with regularization behavior and optimization procedures
- Group Lasso
- Sparse group Lasso
- Fused Lasso
- Hierarchical tree sparsity as an example of structured sparsity penalization
- Trace-norm penalization
- Comparison to horseshoe priors in Bayesian regression

Required reading:

- Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press. (*chapters 1 & 2*)
- Bzdok, D., Varoquaux, G., Grisel, O., Eickenberg, M., Poupon, C., Thirion, B., 2016. Formal models of the network co-occurrence underlying mental operations. PLoS Comput Biol, DOI: 10.1371/journal.pcbi.1004994.
- Forgetta, Vincenzo, Julyan Keller-Baruch, Marie Forest, Audrey Durand, Sahir Bhatnagar, John Kemp, John A. Morris et al. "Machine Learning to Predict Osteoporotic Fracture Risk from Genotypes." *bioRxiv* (2018): 413716.

Week 5: Bias-variance tradeoff

- Linear regression versus k-nearest neighbors
- Comparison of their behavior in high dimensional datasets
- Computability

Required reading:

- Balsubramani, A., Dasgupta, S., Freund, Y., & Moran, S. (2019). An adaptive nearest neighbor rule for classification. arXiv preprint arXiv:1905.12717.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Heidelberg, Germany. (*chapter 1+2*)
- Bzdok, D., Krzywinski, M., Altman, N., 2018. Machine learning: supervised methods. Nature Methods 15, 5-6.

Week 6: Cross-validation

- Why more relevant in large-n datasets?
- Procedure
- Purpose (model estimation/selection/assessment)
- Findings from empirical simulation studies
- What is the best in k-fold?
- Restricted model spaces in model regularization

Required reading:

- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Heidelberg, Germany. (*chapters 7 & 8*)
- Breiman, L., Spector, P., 1992. Submodel selection and evaluation in regression. The X-random case. International statistical review/revue internationale de Statistique, 291-319.
- Varoquaux, G., 2017. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage.

Week 7: Clustering and matrix decomposition techniques

- Structure discovery versus structure inference
- Implicit assumptions of clustering and latent factor procedures
- Why is it ill-posed to search for a best clustering algorithm or clustering solution?
- Why is there no statistical significance of cluster and matrix factorization solutions?
- Cluster validity criteria

Required reading:

- Kleinberg, J., 2002. An impossibility theorem for clustering. NIPS 15, 463-470.
- Eickhoff, S.B., Thirion, B., Varoquaux, G., Bzdok, D., 2015. Connectivity-based parcellation: Critique and implications. Hum Brain Mapp 36, 4771-4792.
- Zamberlan, F., Sanz, C., Martínez Vivot, R., Pallavicini, C., Erowid, E., & Tagliazucchi, E. (2018). The varieties of the psychedelic experience: a preliminary study of the association between the reported subjective effects and the binding affinity profiles of substituted phenethylamines and tryptamines. *Frontiers*, *12*, 54.

Week 8: Diagnostics

- Precision/recall as an alternative to sensitivity/specificity
- Confusion matrix
- F1-score
- ROC curves versus precision/recall curves
- Sample complexity via learning curves in high-bias and high-variance models
- Model calibration

Required reading:

- Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: the Achilles heel of predictive analytics. BMC medicine, 17(1), 1-7.
- Bzdok, D., 2017. Classical statistics and statistical learning in imaging neuroscience. Front Neurosci 11, 543.
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Etemadi, M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94.

Week 9: Statistical Learning Theory

- Important theoretical achievements in classical statistics
- Finite sample theorems
- Hypothesis set
- Approximation error versus estimation error
- VC dimensions
- The fundamental theorem of statistical learning
- Notions of learnability
- Probably Approximate Correct (PAC) Learning
- Uniform convergence

- Non-uniform learnability
- Consistency
- No-free-lunch theorem

Required reading:
  - Shalev-Shwartz, S., Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge University Press. (*chapter 2*)
  - Mohri, M., Talwalkar, A., Rostamizadeh, A., 2012. Foundations of machine learning. MIT Press Cambridge, MA. (*chapter 2*)
  - http://www.no-free-lunch.org/

Week 10: Deep Learning

- Historical overview
- Special empirical properties of deep neural networks (DNNs)
- Autoencoders à la Hinton 2006 Science, why now?
- Types of non-linearity
- Dropout
- Gradient clipping
- Convolutional layers and max-pooling
- Recurrent neural networks
- Long-short term memory neural networks

Required reading:
  - Goodfellow, I.J., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press, USA. (*chapters 1 & 6*)
  - Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504-507.
  - Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering 2, 158 (2018).

Week 11: Model scaling with ever deeper and wider biomedical data

- Bayesianism versus frequentism
- Parametric versus non-parametric models
- Inference versus prediction
- Discriminative versus generative models

Required reading:
  - James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. Springer. (*chapter 2*)
  - Ng, A.Y., Jordan, M.I., 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. Advances in neural information processing systems 14, 841.
  - Bzdok, D., Yeo, B.T.T., 2017. Inference in the age of big data: Future perspectives on neuroscience. Neuroimage 155, 549-564.
  - Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*.

Week 12: Bayesian hierarchical modelling

- Random effects vs. fixed effects
- Sharing statistical strength by partial pooling between parameters
- Correction for multiple comparisons in the Bayesian context

- Highest posterior density intervals (HPDI)
- Markov Chain Monte Carlo (MCMC) sampling
- Posterior predictive checks vs. types of out-of-sample prediction
- Varying-intercept and varying-slope models
- Bayesian ANOVA

Required reading:

- McElreath, R., 2015. Statistical Rethinking. Chapman & Hall/CRC, Boca Raton, FL, USA. (*chapter 13 / multilevel models*)
- Gelman, A., 2005. Analysis of variance—why it is more important than ever. The Annals of Statistics 33, 1-53.
- Bzdok, D., Floris, D.L., Marquand, A.F., 2020. Analyzing Brain Networks in Population Neuroscience: A Case for the Bayesian Philosophy. Philosophical Transactions of the Royal Society B: Biological Sciences.
- Belin, T. R., & Rubin, D. B. (1995). The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statistics in medicine*, *14*(8), 747-768.

Week 13: Reduced-rank regression and other variants of supervised structure discovery in biomedical data

- Canonical correlation analysis (CCA) versus partial-least squares (PLS)
- Principal component regression
- Multi-task learning
- Neighbourhood-component analysis
- Semi-supervised autoencoder

Required reading:

- Cheung, B., Livezey, J.A., Bansal, A.K., Olshausen, B.A., 2014. Discovering Hidden Factors of Variation in Deep Networks. arXiv preprint arXiv:1412.6583.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Heidelberg, Germany. (*chapter 3*)
- Roweis, S., Hinton, G., Salakhutdinov, R., 2004. Neighbourhood component analysis. Adv. Neural Inf. Process. Syst.(NIPS) 17, 513-520.

Assignments and Evaluations

Midterm after week 9*: 30%*
Closed book format, in which objective-type questions will be asked, as well as multiple-choice, and opinion-type questions.

Solutions for coding challenges: 40%
Each week a new coding task will be given that requires to design and implement a data analysis workflow in Python code (except week 13). The correctness of the programmed solutions from students will be checked and scored.

Final Exam: 30%
Closed book format, in which objective-type questions will be asked, as well as multiple-choice, and opinion-type questions. Covers full material.

McGill Policy Statements

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Conformément à la Charte des droits de l'étudiant de l'Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté.

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures" (see www.mcgill.ca/students/srr/honest/ for more information).

L'université McGill attache une haute importance à l'honnêteté académique. Il incombe par conséquent à tous les étudiants de comprendre ce que l'on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l'étudiant et des procédures disciplinaires (pour de plus amples renseignements, veuillez consulter le site www.mcgill.ca/students/srr/honest/).