Machine Learning 2023

Faculty: David Wishart, Vasu Gautam, Mark Berjanskii, Sagan Girod, Michelle Brazas, Nia Hug

August 16, 2023 - August 17, 2023

Contents

Part I

Introduction

Chapter 1

Workshop Info

Welcome to the 2023 Machine Learning Canadian Bioinformatics Workshop webpage!

1.1 Class Photo

1.2 Schedule

Time (Eastern)	**Wednesday, August 16**	**Time (Eastern)**	**Thursday, August 17**
9:45	Virtual Arrivals	9:45	Virtual Arrivals
10:00	Welcome and Technology Check (Nia Hughes)	10:00	Module 5: Gene Prediction with NNs (Lecture and Lab)
10:45	Module 1: Introduction to Machine Learning (Lecture)	12:00	Break (30min)
12:15	Break (30min)	12:30	Module 6: Machine Learning with Keras and Scikit-learn (Lecture and Lab)
12:45	Module 2: Decision Trees (Lecture and Lab)	14:00	Break (1hr)
14:15	Break (45min)	15:00	Module 7 (Continued): Machine Learning with Keras and Scikit-Learn

Time (Eastern)	**Wednesday, August 16**	**Time (Eastern)**	**Thursday, August 17**
15:00	Module 3: Neural Networks (Lecture and Lab)	16:00	Break (30min)
16:30	Break (30min)	16:30	Module 8: Information Extraction with ChatGPT (Lecture and Lab)
17:00	Module 4: Neural Networks for Secondary Structure (Lecture and Homework)	17:45	Survey and Closing Remarks
18:00	End of Day 1	18:00	End of Day 2

1.3 Pre-work

You can find your pre-work here.

Chapter 2

Meet Your Faculty

2.0.0.1 David Wishart

Distinguished University of Biological Sciences and Computing Science University of Alberta Edmonton, AB, CA

— dwishart@ualberta.ca www.wishartlab.com

David was one of the co-founders of the CBW in 1998. He has active research interests in the application of machine learning to a wide range of computational biology problems, from genomics to proteomics to metabolomics. In addition to running a large computational biology lab, David also operates a large wet lab (analytical chemistry, molecular biology, nanotechnology, cell biology, structural biology) supported by its own fabrication shop and electrical engineering group.

2.0.0.2 Vasu Gautam

Senior Scientist Wishart Lab, University of Alberta Edmonton, AB, CA

— vasuk@ualberta.ca

Dr. Vasu Gautam is the senior scientist and Bioinformatics Manager at Wishart Node, University of Alberta. Vasu is intrigued by the diverse world of "omics" and their combined role in biological research, be it proteomics, genomics, or metabolomics. Vasu has worked in both academia and industry in the field of multi-omics. His interest has been to explore the different aspects of these "omics" groups and then combine this knowledge pool to address some of the most difficult questions in the field. Bioinformatics/computational biology has been a great tool in enhancing this capability and continuing his research. His current focus is the study of machine learning algorithms and their applications in different areas of metabolomics.

2.0.0.3 Mark Berjanskii

Research Associate Department of the Biological Sciences University of Alberta Edmonton, AB, CA

— mb1@ualberta.ca

Mark obtained a Ph.D. in Biochemistry at University of Missouri-Columbia, USA. Mark's research interests include NMR metabolomics, bioinformatics, protein NMR structure determination, interactions, misfolding, and dynamics. Between 1996 and 2004, he worked as a member of several research teams that studied the Hsp70-Hsp40 chaperone system. Mark joined Dr. Wishart's group at University of Alberta in 2004. Between 2005 and 2013, Mark studied prion proteins that, when misfolded, cause Mad Cow Disease in cattle and similar Creutzfeldt-Jakob Disease in humans. Since 2004, he has been involved in developing several programs for analysis of protein structure and dynamics, such as Random Coil Index, Preditor, GeNMR, CS23D, PROSESS, Resolution-by-proxy, and Gamdy, as well as bioinformatic analysis and NMR metabolomics.

2.0.0.4 Sagan Girod

Junior Data Scientist Wishart Lab, University of Alberta Edmonton, AB, CA

— lgirod@ualberta.ca

Sagan is a data scientist at the Wishart Node, University of Alberta. His research is based around machine learning and multi-omics with a focus on data curation and database design.

2.0.0.5 Michelle Brazas

Acting Scientific Director Canadian Bioinformatics Workshops (CBW) Toronto, ON, CA

— support@bioinformatics.ca

Dr. Michelle Brazas is the Associate Director for Adaptive Oncology at the Ontario Institute for Cancer Research (OICR), and acting Scientific Director at Bioinformatics.ca. Previously, Dr. Brazas was the Program Manager for Bioinformatics.ca and a faculty member in Biotechnology at BCIT. Michelle co-founded and runs the Toronto Bioinformatics User Group (TorBUG) now in its 11th season, and plays an active role in the International Society of Computational Biology where she sits on the Board of Directors and Executive Board.

2.0.0.6 Nia Hughes

Program Manager, Bioinformatics.ca Ontario Institute for Cancer Research Toronto, ON, Canada

— nia.hughes@oicr.on.ca

Nia is the Program Manager for Bioinformatics.ca, where she coordinates the Canadian Bioinformatics Workshop Series. Prior to starting at OICR, she completed her M.Sc. in Bioinformatics from the University of Guelph in 2020 before working there as a bioinformatician studying epigenetic and transcriptomic patterns across maize varieties.

Chapter 3

Data

3.0.0.1 Course data downloads

Download the scripts and data for this course here

Part II

Modules

Chapter 4

Module 1: Introduction to Machine Learning

4.1 Lecture

Chapter 5

Module 2: Decision Trees

5.1 Lecture

Chapter 6

Module 3: Neural Networks

6.1 Lecture

Chapter 7

Module 4: Neural Networks for Secondary Structure

7.1 Lecture

Chapter 8

Module 5: Gene Finding with NNs

8.1 Lecture

Chapter 9

Module 6: Machine Learning with Keras and Scikit-Learn

9.1 Lecture

Chapter 10

Module 7: Machine Learning with Keras and Scikit-Learn Continued

10.1 Lecture

Chapter 11

Module 8: Information Extraction with ChatGPT

11.1 Lecture