

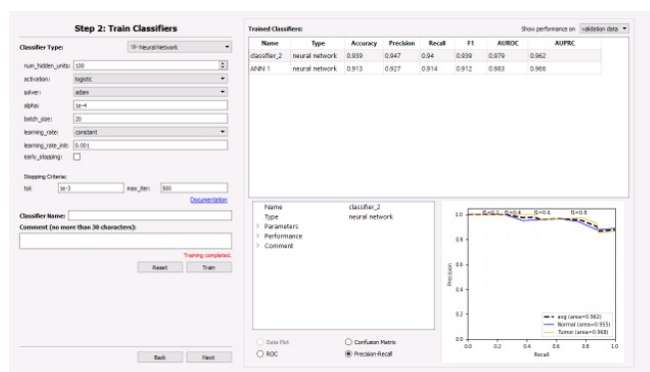


Community Spotlight

Each [Community Spotlight](#) features an outstanding group, partner, resource, or member of our community.

The ml4bio Workshop: Machine Learning Literacy for Biologists

By Fangzhou Mu, Chris Magnano, Debora Treu, and Anthony Gitter



Module Description:

This week's featured resource is a presentation from a [Special Session on Bioinformatics Education](#) at the [2019 Great Lakes Bioinformatics Conference](#) that describes workshop materials, including software with a graphical interface and interactive exercises, for teaching machine learning literacy to biologists. Please find the presentation abstract below.

Machine learning has been incredibly successful in mining large-scale biological datasets. Despite its popularity among computational researchers, machine learning remains elusive to experimental biologists, who form the majority of the life sciences research community, leaving powerful computational tools underappreciated and data generated in wet labs underexplored. Recent years have seen a growing interest among biology trainees to embark on machine learning projects that complement their research. However, most machine learning courses and tutorials require substantial background knowledge in coding and mathematics, which many biologists may lack. On the other hand, bioinformatics workshops for biologists assume less coding experience, but participants are often taught to mechanically run through a software pipeline for certain tasks without learning the best practices in various stages of the workflow. Such an approach, though effective in the short term, can lead to error-prone data analysis, misinterpretation of results, and difficulty in adapting to other tasks in the long run of a scientist's research effort. The community clearly needs to explore novel educational frameworks in order to address these challenges in teaching machine learning to biologists.

Unlike traditional task-centric approaches, our educational objective is to equip biologists with the proper mindset when it comes to applying machine learning in their research and the ability to critically analyze machine learning applications in their domain. Built around this core idea, our ml4bio workshop prioritizes teaching machine learning literacy, that is, the right way to set up

learning problems, how to reason about learning algorithms, and how to assess learned models. We have developed interactive software with a graphical interface and a set of accompanying slides and tutorials for use during workshop sessions. The software and interactive exercises guide participants through a full cycle of the machine learning workflow while doing proper model training, validation, selection, and testing. By following instructions in the slides and tutorials, participants build intuition about the strengths and weaknesses of various model classes and evaluation metrics by visualizing model behavior under different data distributions and sets of model hyperparameters. We further attempt to mind the gap between theory and practice through illustration of machine learning applications on real biological tasks. Overall, our approach encourages beginners to take a holistic view of the machine learning workflow rather than immediately dive into the technicalities of coding and mathematics. We have successfully offered two pilot workshops attended by graduate students and postdocs with diverse backgrounds and research interests. The feedback we collected provides strong preliminary evidence on the effectiveness of our approach.

Moving forward, our short-term plan is to tailor the workshop material to better serve our educational objective and the needs of participants. The current version of the software only supports classification models. For future releases, we will expand the set of models to include those for regression and clustering. We are also looking for new biological case studies that highlight good and bad practices of machine learning in the biological literature. Our long-term software development plan is to more closely link the ml4bio graphical interface and the Python scikit-learn code on which it is built in order to guide participants who wish to later customize their own machine learning pipeline. Our ultimate goal is the national distribution of the workshop. As an initial step towards this end, we are working closely with educators and facilitators on and off campus to outline a timetable on future workshop development and to adopt best practices of successful workshops such as Software and Data Carpentry. Our workshop materials are available at <https://github.com/gitter-lab/ml-bio-workshop/> under the CC-BY-4.0 license and our ml4bio software is available at <https://github.com/gitter-lab/ml4bio/> and [PyPI](#) under the MIT license.

Teaching Setting:

The 3-4 hour workshop described in this resource is suitable for biologists who are interested in learning about machine learning and its application to biological research. Coding experience is not required.

Citation:

Mu, F., Magnano, C., Treu, D., Gitter, A. (2019). [The ml4bio Workshop: Machine Learning Literacy for Biologists](#). GLBIO2019 Special Session on Bioinformatics Education, QUBES Educational Resources. [doi:10.25334/Q44Q97](https://doi.org/10.25334/Q44Q97)

Visit Resource



Share



Tweet

Related Materials and Opportunities:

This resource was presented at a [Special Session on Bioinformatics Education](#) at the [2019 Great Lakes Bioinformatics Conference](#), which was held in May 2019 at the University of Wisconsin at Madison. GLBIO is organized by the [Great Lakes Bioinformatics Consortium](#), an [International Society for Computational Biology \(ISCB\)](#) Regional Affiliate Society, and serves as the primary forum for bioinformatics work in the Great Lakes Region. GLBIO has a long tradition of hosting talks, outreach activities, and interactive workshops in bioinformatics education drawing on a vibrant community of educators from the Great Lakes region and beyond. This year, the program for the first time devoted a [special session](#) specifically to bioinformatics education. You are encouraged to browse abstracts and presentation materials for [invited presentations](#) and [contributed talks](#) from this special session, including [one that was previously featured as a Resource of the Week](#).

The authors of this resource are actively working on revising their workshop materials based on their GLBIO special session and pilot workshop feedback; therefore, you are encouraged to [watch this resource](#) so you'll be notified when a new version of this resource is released.

QUBES on Social Media



BioQUEST is a transformative, collaborative community empowering educators to drive innovation in STEM education for all students.

Copyright © 2024 QUBES, All rights reserved.

P.O. Box 1452, Raymond, NH 03077

You are receiving this email because you have shown interest in receiving updates from BioQUEST and QUBES.

[Subscribe / Unsubscribe](#) from mailing list

[View Community Spotlight on QUBESHub](#)

Community Spotlight: Issue 53