Teaching ML in 2021
An Overview and Introduction

Oliver Guhr [*1] **Katherine M. Kinnaird** [*2] **Peter Steinbach** [*3]

Abstract

With the second edition of Teaching ML, the continuous interest of the community to expand, endorse and validate teaching methods in Machine Learning related fields is strongly underlined. In this paper, we aspire to summarize the annual workshop in order to provide interested readers a head-start. Furthermore, this article is meant as an introduction to the proceedings of our workshop.

1. Motivation

Many experts and practitioners who develop Machine Learning models or infrastructure around these models are confronted with the opportunity to teach Machine Learning at some point in their career. Traditionally, many rely on their gut feeling to design courses that are motivated by these circumstances. The methods of choice are often PowerPoint or similar technologies to present content and `jupyter` notebooks to conduct exercises.

This workshop targets those who would like to know, how teachers from around the globe approach teaching Machine Learning: How deep do they dive into the matter? What mental models do they use to visualize concepts? What media is at play in teaching ML by others?

With this workshop, we hope that all participants obtain a better feeling where they stand with their teaching and how they can improve or collaborate with others.

2. Keynotes from Satellite Event

Each keynote highlighted different best practices to make machine learning knowledge more accessible. For the first keynote, Cornelia Gamst introduced the 'AI Campus' - an online learning platform for artificial intelligence. Sorelle Friedler explained how she embedded ethics in her "Data Structures" classes. In the third keynote, Alicia Johnson talked about how she integrated inclusion and accessibility principles in the writing of her textbook on Bayesian statistics.

Cornelia Gamst explained how the 'AI Campus'[1] provides online courses for teaching students machine learning topics. All materials and nano degrees are provided without any fees, which eliminates important entrance barriers. Furthermore, the team introduced a teaching fellowship that supports tutors to address the problem of high dropout rates in open online courses.

Sorelle Friedler explained how she integrated ethical and environmental topics in her course on data structures. She explained that covering these topics helps to broaden participation in computer science. The assignments of her course cover topics like analysing the bias of risk assessment systems and calculating the environmental impact of computing. She stated that students liked these assignments as they feel more relevant. For educators that like to include ethical topics in their courses as well, Friedler provides the "Teaching responsible computing playbook."[2]

In her keynote "Rethinking the dreaded textbook," Alicia Johnson explained how she did not find inclusive teaching materials for her course on Bayesian statistics. Together with Miles Ott and Mine Dogucu, she created the inclusive and accessible textbook "Bayes Rules!" (Johnson et al.). Johnson stated that "inclusion and accessibility require intentionality and accountability." She explained how she set the goals for her book, tracked and measured them.

3. Paper Summary of 2021

This year's workshop witnessed an exceptional growth in paper submissions. Of 20 papers submitted, the reviewers agreed to accept 16. While some of core ideas of these papers overlap, there are exceptions that were spurred by the Sars-Cov-2 pandemic which affected all instructors and

[*]Equal contribution [1]University of Applied Science (HTW) Dresden, Germany [2]Smith College, Northampton, Massachusetts, USA [3]Helmholtz Zentrum Dresden-Rossendorf, Dresden, Germany. Correspondence to: Peter Steinbach <p.steinbach@hzdr.de>, Katherine M. Kinnaird <kkinnaird@smith.edu>, Oliver Guhr <oliver.guhr@htw-dresden.de>.

Proceedings of the 2^{nd} Teaching in Machine Learning Workshop, PMLR, 2021.

[1]https://ki-campus.org/

[2]https://foundation.mozilla.org/en/what-we-fund/awards/teaching-responsible-computing-playbook/

learners. The following section attempts to provide a high level overview of key ideas from the papers.

Self-guided projects as a common foundation for learning A majority of papers, 12/20 (see Kazmi, 2021; Glauner, 2021; Heras, 2021; Acquaviva, 2021; Weerts & Pechenizkiy, 2021; Raschka, 2021; Valdenegro-Toro, 2021; Shouman et al., 2021; Müller et al., 2021; Chin et al., 2021; Palazzo et al., 2021; Brown, 2021) mention or focus on the use of projects as a foundation for learning. In other words, learners (either individually or in groups) are confronted with (often) curated datasets on which they are asked to practice and deploy their newly acquired machine learning skills. With this, authors point to the fact, that machine learning procedures need to be practiced a lot after being taught at the theoretical or conceptual level. (Glauner, 2021) highlights that incorporating projects into a course is a clear benefit that physical workshops and semester courses have over MOOC based approaches.

In multiple papers, authors emphasize that projects need to be close to the learner's domains of interest. The authors highlight that this approach not only reduces cognitive load but also increases learner engagement. This stands in contrast to standard example datasets like MNIST (Deng, 2012) which stand somewhat remote from learners everyday life perception. (Palazzo et al., 2021) and (Heras, 2021) implement this idea by collaborating with local authorities to obtain locally-based datasets. Further examples include (Acquaviva, 2021) implementing projects aligned with physics use cases, (Kazmi, 2021) using with end-to-end projects for energy engineers, (Müller et al., 2021) delivering content tailored for biologists, (Weerts & Pechenizkiy, 2021) and (Chin et al., 2021) gearing examples toward engineering students.

(Raschka, 2021) takes the idea of self-driven student projects even further and culminate this activity in a peer-review style project presentation in order for the students to experience academic standard procedures. In the same line of thought, approaches like (see Acquaviva, 2021; Canziani, 2021; Eaton, 2021; Chin et al., 2021) also include an activity to discuss published Machine Learning papers to train students in using technical as well as community terminology.

Guidance on structuring a course Another stream of discussion in the accepted TeachingML literature spans the wide field on how to set up a course in terms of technology and approach. Many articles conduct their classes openly on github. This enables both transparency for materials and acts as a facilitator for project management, code access and discussion. Further, the use of literate programming technology like `jupyter` notebooks is encountered in a strong majority of articles. Of exemplary mention, (Canziani, 2021) presents software tools that can be used to prepare remote-only courses including estimates for material preparation by instructors. (Brown, 2021) summarizes how a structured approach to align exercises and examinations with learning goals can also help guide instructor activity as well as learner expectation management. (Marx et al., 2021) add a enlightening turn to the workshops discussion as they present one (of many) approaches to teach machine learning without a computer. This is an inspiring continuation of a theme picked up with (Huppenkothen & Eadie, 2021).

Diversity and inclusion as something for us all Like the keynotes (section 2), diversity and inclusion in machine learning education was also a theme among the accepted papers. This is reflected by contributions such as (Palazzo et al., 2021) who translated existing material to Spanish in order to lower cognitive load for learners. (see Weerts & Pechenizkiy, 2021; van Strien et al., 2021) focus on the aspect of fairness and inclusion also with respect to the training and use of machine learning systems in order to bring this aspect to the learners' attention.

4. Discussions during 2021 Workshop

The Teaching ML workshop centered around building connections and growing the community of machine learning researchers, instructors, practitioners, and educators interested in developing teaching methods specifically tailored for teaching machine learning. Our workshop had two sessions for these kinds of community growing efforts: *Community Connect* and Workshop discussions. The *Community Connect* was like a poster session where authors and participants were encouraged to discuss the accepted short papers. Then later in the workshop, the discussions were more focused conversation about a particular element of teaching machine learning.

The workshop session was set-up in the style of an "unconference" with discussion topics crowd-sourced from participants instead of being dictated by the workshop organizers. The list of topics (in order votes) were:

- ML and Pedagogy, Inquiry-based learning
- Textbooks for teaching ML
- Teaching ML without pre-requisites
- Teaching ML at the undergraduate level
- Teaching ML in STEM (Beyond CS/Math)
- Teaching ML and the Internet of things
- Ungrading (Blum & Kohn, 2020) and/or standards based grading methods

Of these topics, during the discussions two topics garnered the most discussion notes: 1) ML and Pedagogy, Inquiry-based learning and 2) Teaching ML in STEM (Beyond CS/Math).

In the first group – ML and Pedagogy, Inquiry-based learning – much of the conversation concerned scaffolding activities. The theme of using more structured assignments at first and then progressively moving towards more open-ended assignments. There were examples of using sequences of activities like stages of a project or iterative brainstorming tasks to help students craft and execute larger projects. This group also discussed incorporating different levels Bloom's Taxonomy into course learning goals.

The second group – Teaching ML in STEM (Beyond CS/Math) – found commonalities in their diverse teaching experiences (that ranged from more typical lecture/exam systems to those driven by hands-on activities like coding in `jupyter` notebooks). This group noted good practices should include exciting realistic examples early on in a course and remembering that 'easy' problems are not always easy. The group also discussed two concrete ideas to try out in courses: 1) incorporating automatic grading like `nbgrader`[3] and `cocalc`[4], and 2) varying the k in kNN classification to demonstrate limits of algorithms as well as illustrating extreme cases.

5. Conclusion

Future avenues for research This year's submissions yield a lot of course curricula that present introductory material of machine learning to learners. This is inline with the previous iteration (Steinbach et al., 2021). Lowering the entry barrier for learners by lowering cognitive load and involving practicals is a strong component therein. In addition, openness not only with respect to diversity represents another motif that we observed.

Based on this, we can identify at least three major directions where further consideration is needed, including

- Methods for teaching advanced concepts of machine learning
- Quantitatively studying curricula presented so far to infer if the effect on learning outcomes agrees with expectations and feedback
- Investigating ways to convey machine learning concepts without the use of a computer or code

This second TeachingML workshop brought together more teacher-scholars, sharing new ideas for teaching machine learning both in and beyond the 'typical' machine learning fields. We look forward to seeing the community continue to grow in the coming years.

[3]https://nbgrader.readthedocs.io/en/stable/

[4]https://cocalc.com/features/teaching

Acknowledgements

The authors thank the TeachingML 2021 reviewers for supporting the scientific program: Claudia A. Engel, Melody Su, Carola Gajek, Lukas Heinrich, Colin Sauze, David Rousseau, Sebastian Starke, Steve Schmerler, Ludwig Bothmann, John Laudun, Sinead Williamson, Patrick Glauner, Jonathan Heras, Matias Valdenegro, Mike Trizna, Guillaume Muller, Martin Palazzo, Thiemo Leonhardt.

Katherine M. Kinnaird is the Clare Boothe Luce Assistant Professor of Computer Science and Statistical and Data Science at Smith College and as such, is supported by Henry Luce Foundation's Clare Boothe Luce Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Luce Foundation.

Peter Steinbach was funded by Helmholtz Association's Initiative and Networking Fund through Helmholtz AI.

Oliver Guhr has been funded by the European Social Fund (ESF), SAB grant number 100339497 and the European Regional Development Funds (ERDF) (ERDF-100346119).

References

Acquaviva, V. Teaching machine learning for the physical sciences:a summary of lessons learned and challenges. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=z-N7JMHjLO.

Blum, S. D. and Kohn, A. *Ungrading: Why rating students undermines learning (and what to do instead)*. West Virginia University Press, 2020.

Brown, S. M. Participatory live coding and learning-centered assessment in programming for data science. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=q8N3WvWvt4X.

Canziani, A. Teaching deep learning, a boisterous ever-evolving field. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=3jKiduENcO.

Chin, R., Stamoulis, D., and Marculescu, D. Putting the "machine" back in machine learning for engineering students. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=dNh4TiaOLy_.

Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

Eaton, C. D. Teaching machine learning in the context of critical quantitative information literacy. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=k9jaVBHot_.

Glauner, P. Staying ahead in the MOOC-era by teaching innovative AI courses. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=4MLUG42Q6Xh.

Heras, J. Deep learning projects from a regional council: An experience report. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=Z1kgcLVfha.

Huppenkothen, D. and Eadie, G. Teaching the foundations of machine learning with candy. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 29–35. PMLR, 2021.

Johnson, A. A., Ott, M. Q., and Dogucu, M. *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. Chapman and Hall/CRC. URL https://www.bayesrulesbook.com/.

Kazmi, H. Teaching machine learning through end-to-end decision making. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=4rOjJXIcU0.

Marx, E., Leonhardt, T., Baberowski, D., and Bergner, N. Using matchboxes to teach the basics of machine learning: an analysis of (possible) misconceptions. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=huJogZLN2t.

Müller, R., Fasemore, A. M., Elhossary, M., and Foerstner, K. U. A lesson for teaching fundamental machine learning concepts and skills to molecular biologists. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=knwKgaspObQ.

Palazzo, M., Velazquez, A., Breda, M., Callara, M., and Aguirre, N. Teaching machine learning in argentina: the clusterAI pipeline. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=Qf691VTesDa.

Raschka, S. Deeper learning by doing: Integrating hands-on research projects into a machine learning course. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=yFPqbprG2Qb.

Shouman, O., Fuchs, S., and Wittges, H. Experiences from teaching practical machine learning courses to master's students with mixed backgrounds. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=ijuM-MVwVEk.

Steinbach, P., Seibold, H., and Guhr, O. Teaching machine learning in 2020. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 1–6. PMLR, 2021.

Valdenegro-Toro, M. Teaching uncertainty quantification in machine learning through use cases. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=QoB8QGu5ZSL.

van Strien, D., Bell, M., McGregor, N. R., and Trizna, M. An introduction to AI for GLAM. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=62qGdDGSyqr.

Weerts, H. J. P. and Pechenizkiy, M. Teaching responsible machine learning to engineers. In *ECMLPKDD 2021 Workshop TeachML*, 2021. URL https://openreview.net/forum?id=m28wDC7B3kx.