Built to Last: Lessons on Fostering a Student ML Community

Elizabeth Lau
Machine Learning at Berkeley
lizzielau@berkeley.edu

Valeriy Rotan
Machine Learning at Berkeley
valeriy.rotan@berkeley.edu

Ashwin Reddy
Machine Learning at Berkeley
adreddy@berkeley.edu

Michael Equi
Machine Learning at Berkeley
michaelequi@berkeley.edu

Arjun Sripathy
Machine Learning at Berkeley
arjunsripathy@berkeley.edu

John Ian So
Machine Learning at Berkeley
johnianrso@berkeley.edu

ml.berkeley.edu

Abstract

As the frontiers of machine learning (ML) continue to expand, the gap between the public understanding of ML and state-of-the-art research widens. While laboratory researchers benefit from easily accessible and encouraged collaboration with domain experts, the same cannot be said of newcomers to the field. At the undergraduate level, where socioeconomic inequality means some students have stronger backgrounds than their peers, increasing the accessibility of practical, hands-on opportunities in machine learning is essential to narrowing this gap. In this paper, we detail the approach of Machine Learning at Berkeley (ML@B), a university-based undergraduate student organization aimed at bridging this gap by encouraging collaboration with established figures in the field as well as within the organization itself. We have found integral to this process the practice of placing individuals in roles where they both create value for themselves and others enabling personal motivation to drive group success. This principle has guided the formulation of the diverse set of initiatives and programs we discuss here. We hope that the perspectives gained from ML@B provide insights into successfully integrating undergraduates into a technical environment and fostering an academic culture that encourages collaboration.

1 Introduction

Founded in 2015, Machine Learning at Berkeley (ML@B) is a student-run non-profit ML organization based at the University of California, Berkeley. ML@B's mission is to empower its members within the ML space, allowing them to pursue further opportunities beyond the organization. Collaboration in ML@B can be broadly categorized into *internal collaboration*, which transfers knowledge and experience between generations of the organization via mentorship and technical projects; and *external collaboration*, which introduces novel teachings, opportunities, and connections to the ML@B community and beyond via open source education and partnerships with established organizations. ML@B has found empowering every member to craft their unique role, one that contributes to the learning of others and themselves, creates the optimal environment for collaboration.

Workshop on Broadening Research Collaborations in ML (NeurIPS 2022).

2 Internal Collaboration

Internal collaboration in ML@B encompasses peer-driven efforts to further members' ML knowledge and experience. We outline the organization's approaches below.

2.1 New Member Education Program (NMEP)

The New Member Education Program (NMEP) [6] is a tailored curriculum of hands-on machine learning basics taught by older members and consisting of weekly lectures, assignments, and an open-ended final project deliverable. Most new members who join ML@B go through NMEP, a cornerstone of ML@B's mentorship initiatives since 2018. By the end of the semester, a diverse cohort of ML newcomers is equipped for advanced exploration in ML. At the same time, those teaching benefit from strengthening their fundamentals and improve their ability to communicate challenging technical topics with clarity sufficient for beginners to understand.

Program Structure ML@B implements peer-to-peer instruction as described in [4] and [5] to promote cohesion and plant seeds for future collaborations between new members. Weekly sessions begin with presentation of ML fundamentals before transitioning over the semester to more modern work and highlighting active areas of research. NMEP focuses on creating an interactive and personalized learning environment giving every student a chance for mastery. For example, to promote interactivity, new members present weekly to the rest of their cohort foundational papers relevant to previous lectures creating a student-led discussion. Beyond the classroom, students are expected to complete weekly coding assignments with homework partners and encouraged to attend homework parties to work together and ask questions to older members.

NMEP centers on project-based learning. Students are given complete agency over a NMEP capstone project early in their ML@B careers. Each cohort is broken into small groups and mentored by a senior ML@B member with relevant domain knowledge to the project. Part of the capstone deliverable is a presentation to all of ML@B as an important opportunity to receive feedback on their work and feel more included in the community.

2.2 Fellowship Program

Similar to visiting researchers in a professional lab setting, ML@B's Fellowship program enables more advanced students (*fellows*) with domain-specific expertise to start initiatives that leverage their experience and ML@B's resources. Fellows often come into ML@B as graduate or exchange students with diverse backgrounds, leading to a fresh influx of perspectives for them and existing members. Fellows have the flexibility to mentor projects, present workshops and papers relevant to their past work, push focused research, or innovate entirely new collaborations. As an added benefit, fellows often share their ML@B experiences with other communities they are a part of and draw future fellows to ML@B.

2.3 Academic Culture

ML@B's academic culture aims to equip and encourage members to read, digest, and present technical research material with ease. Inspired by experiences in an actual lab, this aspect of ML@B culture gives members opportunities to grapple with technical topics while simultaneously providing value to an audience. Academic culture is an opportunity that all members are encouraged to leverage from the start of their careers in ML@B; because of this, many rapidly gain the skills and experience needed to excel in further academic settings.

Reading Groups and Workshops Weekly reading groups allow ML@B members to develop a deep understanding of a particular research niche and empower them to introduce others interested to the space. For example, they may technically discuss specific papers or survey a broader range of work. Workshops play a similar role to reading groups, but instead are focused on a specific application or ML engineering tool. Workshops and reading groups aim to add value for both the audience, who get early exposure to a variety of research niches, and the presenters, who can deep dive into their ML curiosities, explore more recent advances in literature, and gain additional insights from a technically diverse audience. Furthermore, even the development of these presentations may

create opportunities for collaboration: if a less experienced member would like to present, another member with the relevant domain knowledge can provide the required mentorship.

Research Meetings ML@B internal research meetings aim to offer a platform for members actively involved in research to present their work in a setting that gives them a greater voice and opportunity than most graduate student dominated external labs. Additionally, it allows students not yet experienced to join future external lab meetings with the ability to carry out research dialogue, developing their academic comprehension and critical thinking abilities.

Wiki The ML@B wiki is a digital library that documents member insights from generation to generation. To the past, present and future ML@B members, the wiki is rich with learning resources such as reading lists, past event recordings, member directories, and ML@B growth.

Advanced Topic Study Groups In advanced topic study groups, members push themselves to deep dive into areas they find challenging. One such example was an *Underactuated Robotics* focus group in the 2021-2022 school year. This self-driven group led weekly meetings consisting of members with varying levels of experience coming together to discuss topics of interest and present on chapters in [8]. In focus groups, members are able to both teach and learn with their peers, resulting in a collaborative environment where everyone feels equal regardless of seniority and previous experiences.

2.4 Social Culture

A core benefit of ML@B over many traditional research environments is the emphasis on peer-to-peer bonding and frequent social interaction. Strong ties within the organization have led to long-term collaboration, which helps encourage the free flow of ideas and a willingness to work together on hard problems. This is emphasized by alumni who stay in touch with the organization and have provided project partnership opportunities with companies and labs they work at. Alumni have also have founded startup teams comprised of fellow ML@B members.

3 External Collaboration

As internal collaboration aims to grow internal knowledge and perspectives within the organization, *external collaboration* aims to provide an interface and dialogue between the organization and the greater ML community. We outline several initiatives below.

3.1 Open Source Education

The common goal of all ML@B's open source education initiatives is to provide resources for underrepresented communities in pursuit of a more diverse ML community. Open source education initiatives at ML@B demand a focused understanding of topics being taught, and provide a large scale audience and experience to those teaching.

Courses ML@B focuses on bridging the gap between underrepresented groups and the ML academic landscape, specifically for underclassmen and gender minorities. To the immediate Berkeley campus, ML@B teaches a public course geared towards underclassmen with no technical or ML background on varying topics such as deep learning for computer vision, self-driving, and ML medical applications. Current and past open source education efforts prioritize making content available in an easy-to-understand manner to lower the bar of entry into ML.

In the past, ML@B has run a public focus group initiative known as the Guided Resource Education Program in ML to diffuse similar NMEP foundations to a wider audience beyond ML@B. ML@B also collaborates with campus groups such as the Association of Women in Electrical Engineering & Computer Science (AWE) and CS Kickstart to establish formal mentorship and various teaching events for gender minorities. As a melting pot of backgrounds, ML@B understands diversity is critical for robust collaborations between all levels of experience.

Blog Blog writing provides an outlet where any ML@B member can reach a wider audience, challenge themselves to produce original opinions, and sharpen their technical writing skills. Blogs

often allow for cross-disciplinary exploration such as [7], an examination into art-generation systems or [3], an overview of ML impact in neuroscience and healthcare. Blog also has the added benefit of building an online community and sparking future collaborations for ML@B.

3.2 Team Projects

Our discussion thus far has focused on the cultivation and dissemination of knowledge within the organization and beyond. Consulting, research, and competition projects play the key role of allowing members to turn this knowledge into meaningful work powered by collaboration of members with varying backgrounds and external guidance. The resulting gradient of experience allows all members the opportunity for technical growth while providing the support necessary for everyone to contribute.

Consulting Projects In consulting projects, members work towards high impact company goals often with a chance to improve core technologies and services. In the past, these collaborations have been effective because companies do not need to build infrastructure for in-house ML research and development; simultaneously ML@B members may explore varied applications of ML and develop industry level work-habits. Often the entities we work with are experts in non-ML fields such as finance or 3D design, producing fruitful cross-disciplinary collaborations.

Research Projects Research projects allow ML@B members to push their own research agenda creating a more accessible preview of what PhD students experience. In doing so they create opportunities for less experienced members, especially those not ready for external labs, to contribute meaningfully to a research project. ML@B connects each research project team with a domain expert in academia or industry capable of providing invaluable guidance. For example, [2] presents work from 4 ML@B members guided by a leader in Adversarial ML working at Google Brain.

Competition Projects ML competitions distill a challenging real world problem into maximizing a clearly defined metric on openly released challenge data. Competing in such events gives teams not only the opportunity to engage in highly purpose driven collaboration internally, but also an avenue to interact with others working on the same problem globally by stimulating an interactive flow of creative research oriented ideas across borders.

3.3 Growing Network

ML@B's belief in the importance of collaboration drives its effort to build a professional network which extends beyond the organization. A few examples include weekly general meetings that host guest speakers in leading research and industry roles, a semiannual research symposium we facilitate along with other undergraduate groups on campus, and an annual Machine Learning Data Science career fair where we connect hundreds of students with companies in the space. Regardless of experience, everyone is able to benefit from and contribute to this network in some way and that allows it to naturally grow and create exciting opportunities for all those involved.

4 Case Study: Language and Decision Making

In this section, we present an ongoing collaboration effort within the organization; internally named *RL-Lang*, this group focuses on how natural language provides useful priors for topics in embodied decision making, such as exploration, hierarchies, and goals.

The initiative initially started as an NMEP capstone project in reinforcement learning (RL). Under the guidance of an older member currently doing research in the area, the group began exploratory work in RL, including re-implementing common algorithms and working with relevant infrastructure. Following the completion of the capstone project, the project converged on language as a specific direction and transformed into an advanced topics focus group over the break, undergoing literature review on ideas in language conditioned imitation learning, representation learning, and general exploration.

At the beginning of the following semester, the topics group was given agency as a project group; older members were brought on as both advisors and senior contributors. In addition to providing mentorship for younger members, older members leveraged their external resources, including

connections to labs and workshops, to solidify directions of progress. At the same time, younger members had plentiful opportunities to work and become comfortable with state-of-the-art algorithms and datasets. The group currently leverages many of ML@B's resources for guidance, including input from guest speakers and alumni from the SayCan team [1].

Collaborations such as these make full use of ML@B's unique resources; many of the external connections brought in by older members for *RL-Lang* have now transformed into explicit partnerships, collaborations, and advisorships.

5 Conclusion

In this perspective paper, we have presented the various opportunities ML@B creates for its members to facilitate and inspire natural collaboration for the development and transfer of ML knowledge in both academic and applied settings. All of the organization's initiatives are designed to foster collaborations which offer all involved members experiences tuned to their depth of knowledge and areas of interest. While the organization's internal collaborations demonstrate the effectiveness of these processes operating among college students, its external collaborations also suggest there are meaningful ways far more senior individuals can participate in the system as well. We are hopeful that the ideas we have shared can help guide budding academic communities that strive for similar goals.

6 Acknowledgements

The authors also wish to recognize and acknowledge the Machine Learning @ Berkeley community as a whole, and several individuals for their mentorship and discussions to our work. We thank Tyler Zhu for his guidance as well as Aryan Goyal, Arvind Rajaraman, Nitish Dashora and Liam Anthony for their helpful observations and insights.

References

[1] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. Do as i can and not as i say: Grounding language in robotic affordances. In *arXiv preprint arXiv:2204.01691*, 2022.

[2] Oliver Bryniarski, Nabeel Hingun, Pedro Pachuca, Vincent Wang, and Nicholas Carlini. Evading adversarial example detection defenses with orthogonal projected gradient descent. In *International Conference on Learning Representations*, 2022.

[3] Ana Cismaru. A survey of applications of ml in healthcare, 2021. https://ml.berkeley.edu/blog/posts/health-ml/.

[4] Adam P. Fagen, Catherine H. Crouch, and Eric Mazur. Peer instruction: Results from a range of classrooms. *The Physics Teacher*, 40(4):206–209, 2002.

[5] Eric Mazur. Farewell, lecture? *Science*, 323(5910):50–51, 2009.

[6] Valeriy Rotan. Ml@b new member education program, 2021. https://ml.berkeley.edu/nmep/.

[7] Charlie Snell. Alien dreams: An emerging art scene, 2021. https://ml.berkeley.edu/blog/posts/clip-art/.

[8] Russ Tedrake. *Underactuated Robotics*. 2022.