

Report on the second hackathon

Lucio Anderlini, *on behalf of the organizers*

ML-INFN meeting, 2022-01-10

Second hackathon

- **Entry level**, again, due to overwhelming applications for the first date
- Virtual only (due pandemic)
- 13-15 Dec 2021

The idea was to replicate the first hackathon:

- keep the programme as similar as possible

The screenshot shows a web browser window displaying the event page for the "Second ML-INFN Hackathon: Starting Level". The page features the ML-INFN logo and the event title. The date is 13-15 Dec 2021, and the location is Zoom. The page includes a navigation menu on the left with options like Overview, Organizing Committee, and Registration. The main content area contains a welcome message, a note that registration is closed, and a list of topics to be covered during the hackathon.

Second ML-INFN Hackathon: Starting Level

13-15 Dec 2021
Zoom
Europe/Rome timezone

Overview

- Organizing Committee
- Timetable
- Contribution List
- My Conference
 - My Contributions
- Registration
- Feedback Surveys
- Pre-tutorial Self-teaching
- Technical prerequisite
- Zoom Videoconferencing
- Groups and Resources for the hackathon
- Telegram Bot and Communications

Welcome to the second edition of the Machine Learning @ INFN (ML-INFN) starting level hackathon, **dedicated to INFN Affiliates**.

Registration is now closed.

You can see the first edition [here](#), with most of the talks attached as video files.

The course is to be considered as "starting level" for Machine Learning topics. The hackathon will be organized over 3 days, distributed as

1. General introduction on ML and on its use in INFN (including Clouds)
2. Tutored hands-on of specific use cases, attempting to reach fully working products; a review of the ML utilization in specific use cases of INFN interest
3. The hackathon, with participants working in groups trying to achieve a goal in the form of a realistic analyses. In the latter part, presentation of their work is expected and discussed among all the groups.

Registrations

We set a participant cap at 60 (10 groups of 6 students) with a ~10% overbooking.

Number of registrations: 63

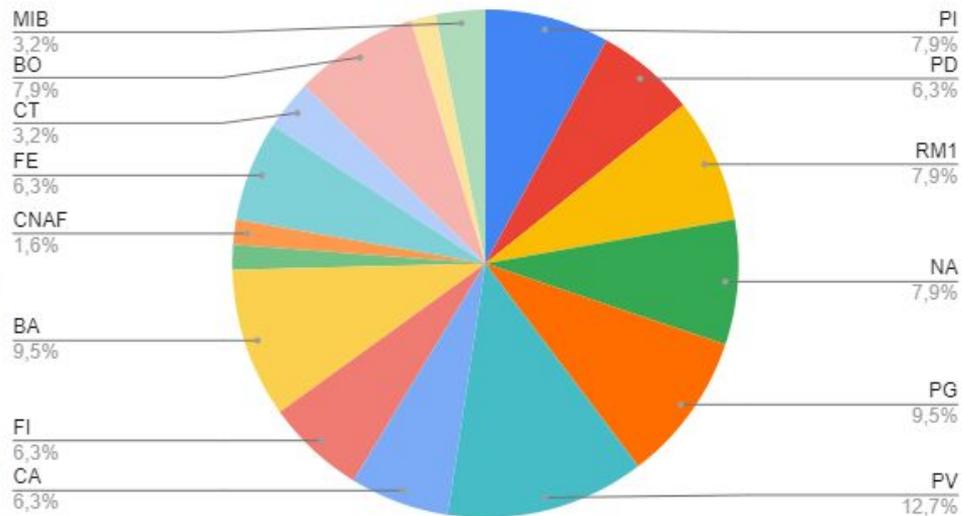
Withdrawals (communicated before the start): 4

Requested topic	First choice	Second option
Segmentation of lesions in medical physics	11	17
Classification @ LHC	33	17
Autoencoders for VIRGO/GW signal analysis	19	29

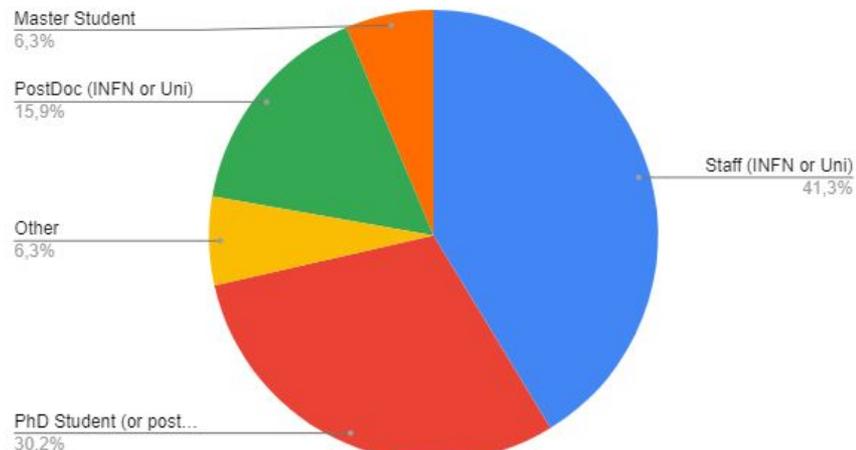
51 participant got their 1st choice satisfied. 6 participant got their 2nd choice.

Some statistics of the participant

INFN Structure



Status



Wide spectrum of background and career levels

Technical setup

- No significant problem observed for registration and login to the INFN Cloud resources
- Excellent jupyterlab environment with important improvement over the previous edition
- Collaborative jupyter has been evaluated still too unstable for 6 concurrent users plus a tutor: *used the same “file-sharing” technique as for the previous edition.*

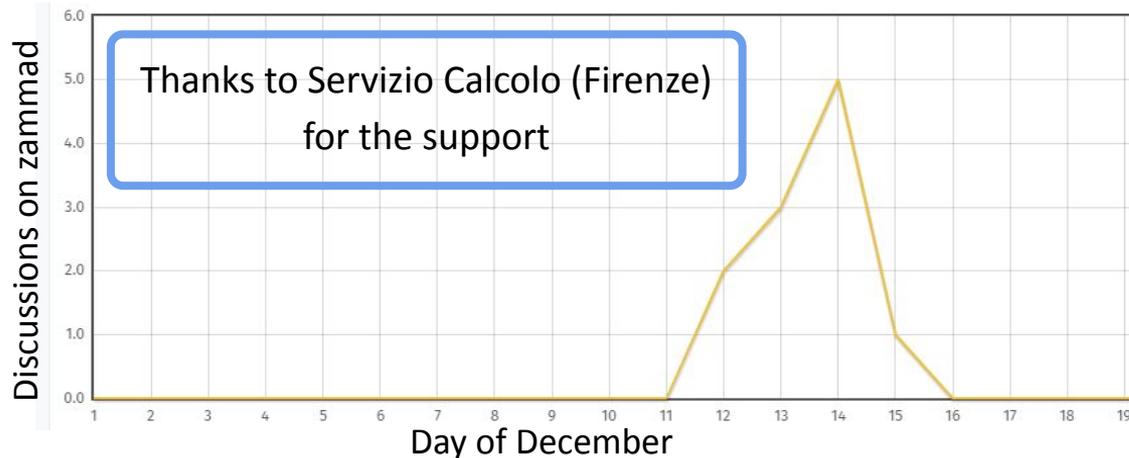
Thanks to the INFN-Cloud team for their crucial support

Questions & Answers *during the first two days*

We received most of the questions via zoom, either in the chat or raising hands.

Interesting discussions happened through zammad.

Mailing list and personal emails were used in few occasions.



Cruscotto

Paroramiche

knowledge base

- Good Morning, I'm not able to c...
- Salve, Ho provato l'accesso a inf...
- hai, I am trying Exercises_empty...
- hello! some generic questions. 1...
- Ho due domande 1) come vengo...**

Gianfranco Paternò

- Grazie mille, ti chiedo solo un ul...
- joining room? crashed ...

GP

Ho due domande 1) come vengono definiti i filtri in un layer Conv2D? Voglio dire, in base a cosa si fissa il primo argomento del metodo?
2) il codice keras viene automaticamente eseguito su GPU in caso ce ne sia una disponibile?

imposta come interno rispondi 14/12/2021 dividi

LA ciao Gianfranco

imposta come interno 14/12/2021 dividi

LA 1) il numero di filtri è un'indicazione della complessità del layer, un po' come il numero di neuroni in un dense layer. Di solito quello che si fa è partire con un numero piccolo di filtri nei primi layer e poi aumentare il numero di filtri. Questo con l'idea che man mano che si procede verso uno stadio più avanzato, la rete impara "concetti" più complessi e a ciascuno di questi concetti viene associato un "filtro". Ad esempio, i primi layer di una CNN per object detection imparano a distinguere linee orizzontali da verticali, poi layer intermedi riconoscono un cerchio da un quadrato e alla fine distinguono un cane da un gatto

imposta come interno 14/12/2021 dividi

LA 2) sì! Purchè l'ambiente sia adeguatamente preparato

imposta come interno 14/12/2021 dividi

LA (ci deve essere cuda e tensorflow per gpu installati, per intendersi)

imposta come interno 14/12/2021 dividi

LA ma dal punto di vista del python non cambia niente.

imposta come interno 14/12/2021 dividi

1) siamo che qualitativamente avevo capito, quello che non capisco è se sotto 5 filtri, come sono definiti? E come si sa che i primi permettono di distinguere linee verticali

Survey response (31/34 responses)

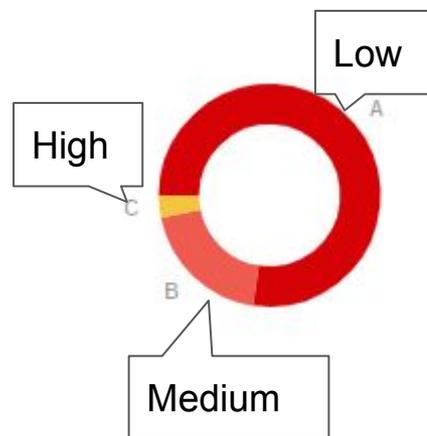
The latest 3 responses are not included in the statistics, only their free-text comments are.

Proficiency before the course

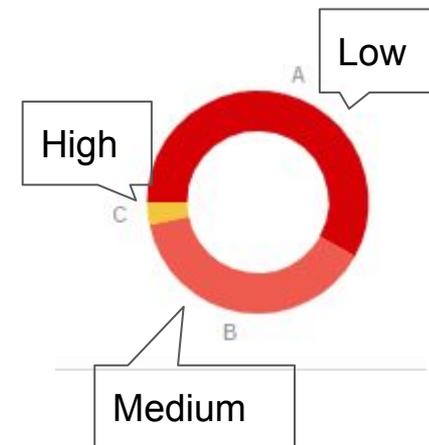
Python



Machine Learning

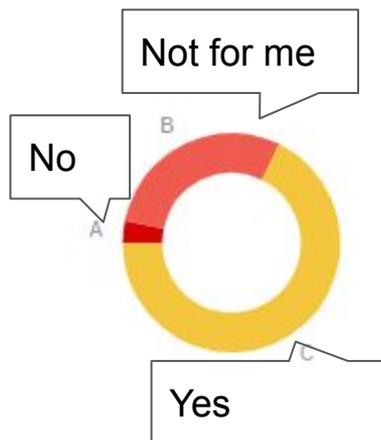


Cloud technologies



Self training

Was self-training useful?

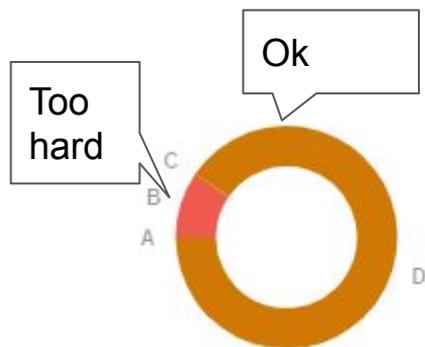


Comments

*Self-training with **numpy** was often **not sufficient** to smoothly follow the hands-on examples*

*if possible, spread lectures and hands-on sessions on the **entire week!** :) many thanks!*

ML introductory lectures

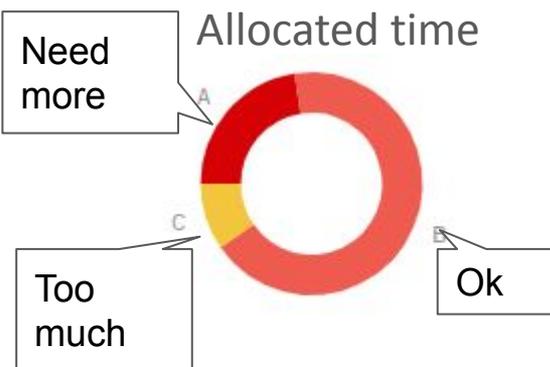


The two lessons explained fundamental concepts which were needed during the rest of the hackaton. **MUCH** more time should be devoted to them. Pierini's lesson was very interesting but **often too technical** to be fully understood. Giagu's talk was better tuned for the level of the audience.

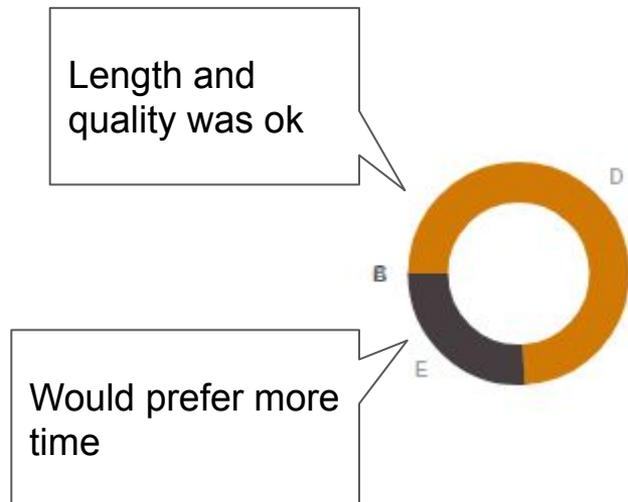
Maybe **alternate theoretical lectures with the hands-on**. Found too much complex the lectures while the hands-on very useful

It would be nice to have at least a **one day break** between the lessons and hands-on session and the final hackathon exercise

Too many information in a too short time. Maybe a short document describing the ML basis should be given to the attendees before the start of the Hackathon. In this way the presentations would be more useful. I have **found myself lost in the many acronyms** I didn't know....



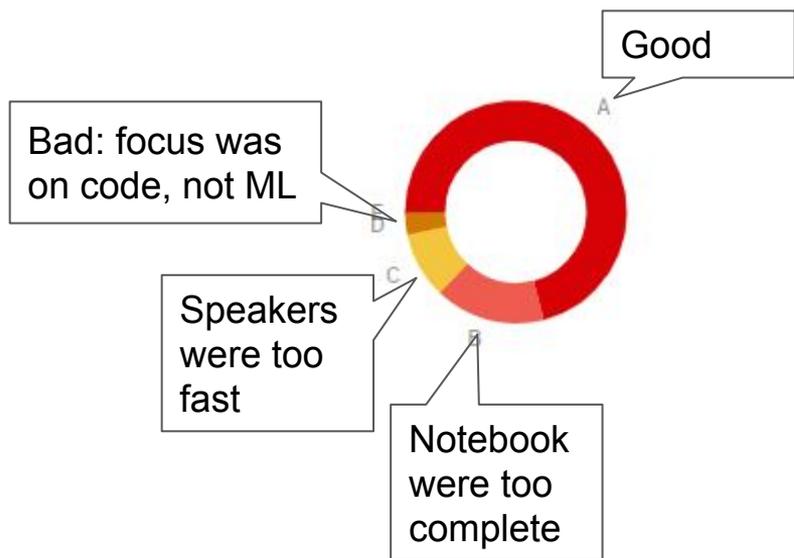
Seminars



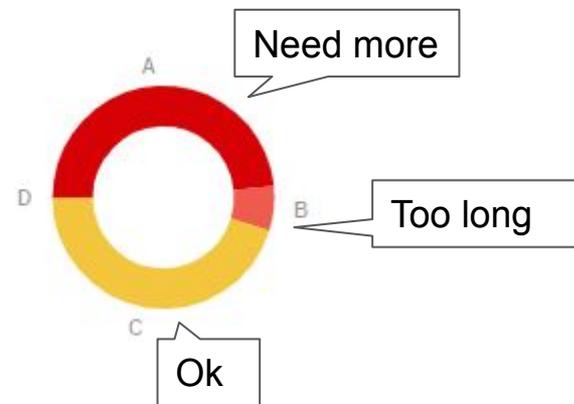
Additional suggested topics:
Dark matter beyond HEP

Hands-on

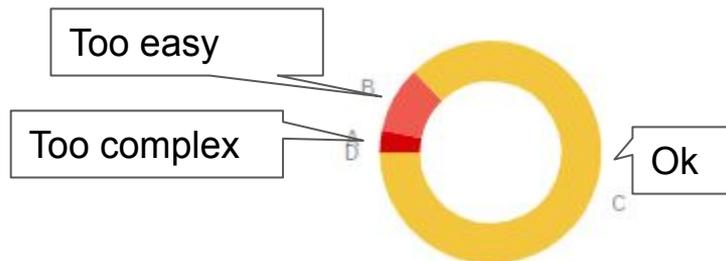
Using notebooks



Allocated time



Difficulty level



Hands-on - Comments

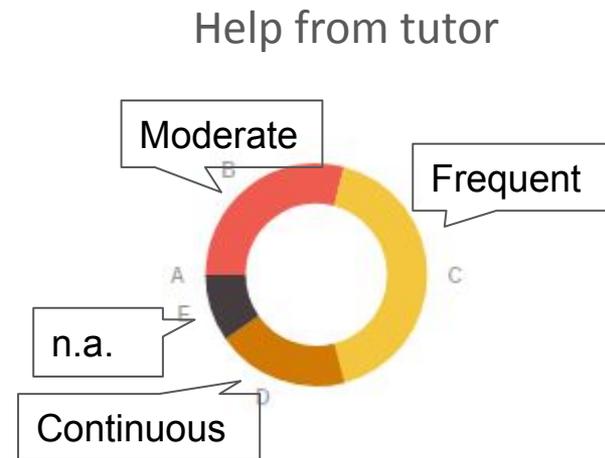
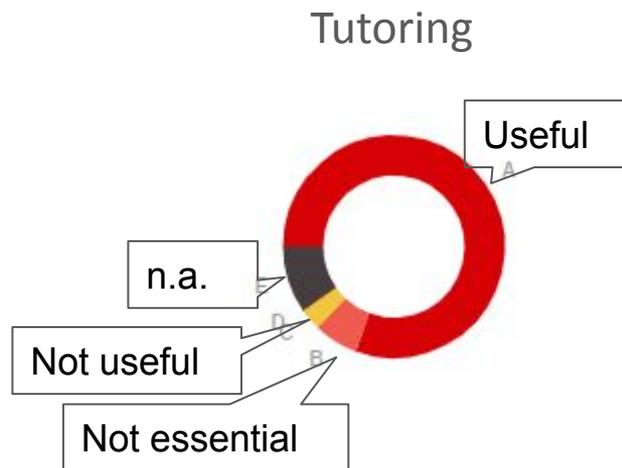
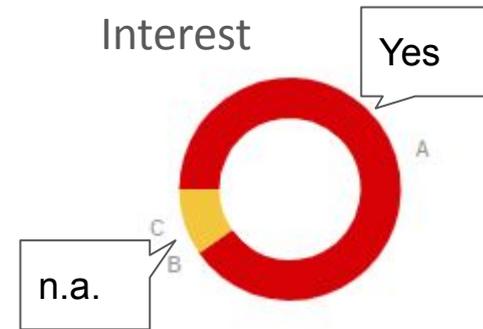
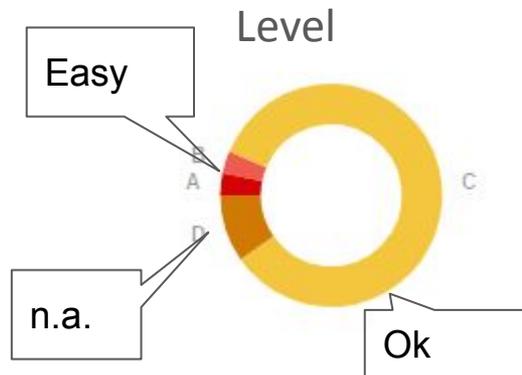
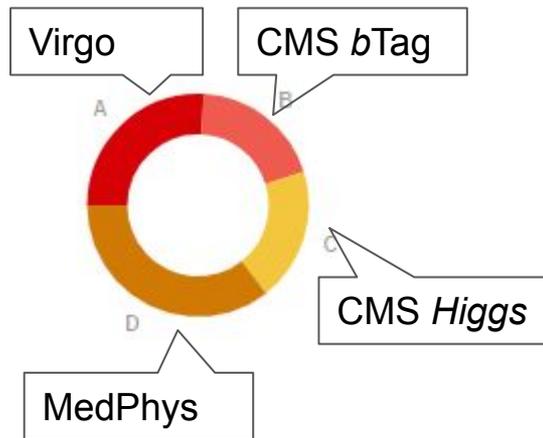
*The amount of information was way too much, the allocated time was too little, **use of supporting packages** (numpy, pandas, etc.) **was too advanced**. I did understand a bit of the first two sessions (Anderlini, Rizzi) but was completely lost with the other three. In my opinion this section of the hackaton should be retought (see below)*

Alessandro Bombini's presentation was too complex for the time allocated

I think that more time spent here would be very useful to get more "take home messages"

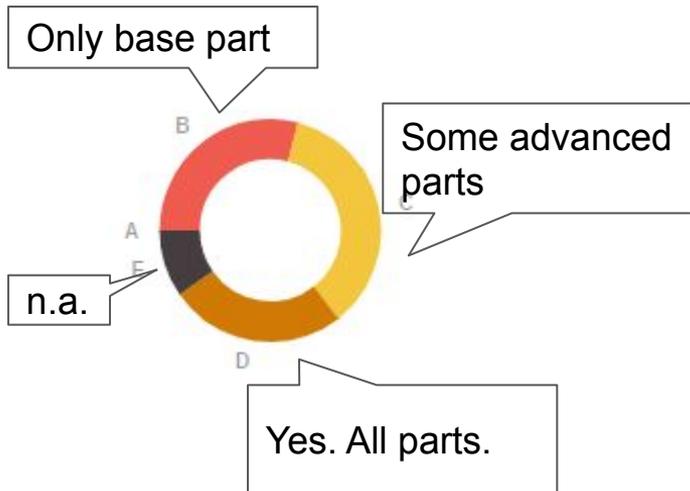
The Hands-on were good enough. Some maybe too easy, some too complex, but overall it was OK.

Hackathon

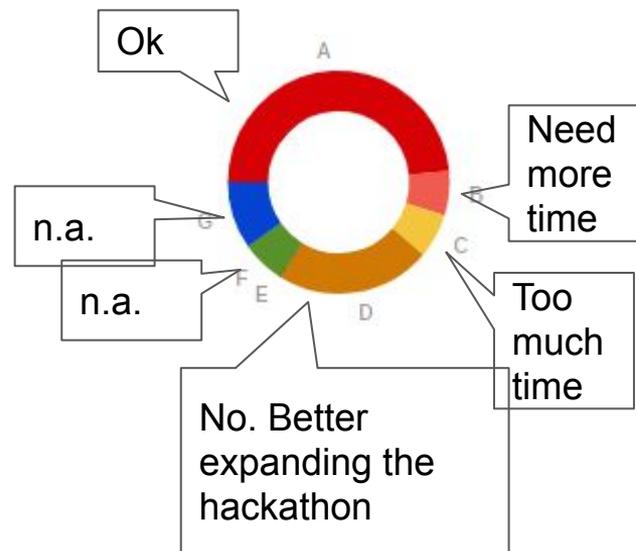


Final reports

Your group completed the hackathon?



Final discussion



Suggestions 1

Requests for more time
Other suggestions

Warnings
Appreciations

*My suggestion is to allocate **more time for the "hackathon"**, for example a **couple of weeks** with a more complex problem, and allocate a couple of half day in which the students can interact with the tutor. (two weeks of work, 1/2 day with the tutor at the beginning, 1/2 day with the tutor after one week, and 1/2 day at the end to present the results of the tests.*

***More time for the hackathon** could have been useful. Maybe some introduction slides for the physics, and removing all the non-code or non-code explanation parts from the jupiter notebook.*

*The actual hackaton was the best part of the course. Thanks to a) the concepts explained the first day, b) the techniques shown in the first two hands-on sessions and c) the continuous presence of the tutor, **I was able to fully understand the technical aspects of the exercise and I am now able to reproduce it on my own. Time allocation for the hackaton should be greatly expanded.***

Suggestions 2

Requests for more time
Other suggestions

Warnings
Appreciations

*The hackathon was great, having a tutor on hand all the time to ask questions/receive feedback was really useful and the interactive nature of working in a group was really helpful too. Writing a short report at the end was helpful but I didn't find it very helpful listening to the reports of other groups. For me, after a *short* introduction to the theory and applications of ML, I think it would have been more helpful to go straight to hackathon sessions. It would have been nice to be able to participate in different groups so that you could get an idea of different types of NNs and the different things they can be used for, especially because image recognition is fairly standard and perhaps easier to self-teach than other types of NNs. I definitely benefitted from the interactive session though, more than the other parts!*

The final discussion needed more time to talk about networks.hackathon would need more time in total.

The time dedicated to the hackathon was too short. In the end the tutor was answering most of the time (in the second part of the hackathon) because the time was running out. More time, less intervention from the tutor.

The tutor was a bit patronizing, but the main issue was that the group was too heterogeneous, with very different python/ML proficiency levels. This made the hackathon messy and complicated for the beginners and extremely boring for those with more experience.

Suggestions 3

Requests for more time
Other suggestions

Warnings
Appreciations

I both enjoyed the course and (I think) benefited from it. Above I noted several problems, especially with the hands-on examples: **at the end of the second day I felt terribly frustrated and was actually tempted of not doing the final hackaton.** A better time allocation would have been: basic concepts and applications the first day (mostly passive learning, more space for the basic concepts), technical examples the second day (at a **slower pace and with less examples**), tutor-intensive hackaton the third day, (almost) **tutor-less hackaton the fourth day.** I think the course would indeed benefit from an extra day. **Thanks to the organizers, the speakers and the tutors for the tremendous job. Please keep me informed of future hackatons.**

This experience was very useful, I already had some basic knowledges on ML techniques but the hackathon helped me to better understand the crucial and fundamental aspects. I believe that the tuesday afternoon, during which we have been left free to "re-study" the several notebooks, it was helpful and maybe more time could be dedicated to this "own preparation " before the real hackathon (maybe **diluted on the three days, that could become four?**). In any case, **I am very sastisfied and enthusiastic about the course, thanks a lot, see you to the next ML-INFN hackathon! :)**

I think the hands-on in group is fine, but there should be a **strong invitation to have every participant to carry out his/her own exercise instead of watching the person in charge for that shift to write the code,** otherwise is too easy, until it's too late

Suggestions 4

Requests for more time
Other suggestions

Warnings
Appreciations

Hopefully INFN will provide common/shared resources with GPU(!) for interactive network development for single researchers who have no other machines available.

It was a fantastic school. If only it lasted longer! E.g. a week.

Arguments are way too complex to shrink the Hackathon to 3 days. In general the HEP part was a little too difficult to follow for a non-HEP background. My suggestion would be to dedicate a school like the Cern Thematic School of Computing.

My general feedback on the ML hackathon is very positive. I found it useful and well organized. There are several starting point to explore and learn more in the talks and notebooks via self teaching. I would have found useful, a few more slides with suggestions to learn deeper. This comment doesn't change the general comment that the hackathon was very good!

Probably adding one day might be "useful" to get more concepts. But it was really really great anyway

I would appreciate a longer exercise (our hands-on) session.

I like it very much. It was very helpful to enter and understand the world of ML. Less theory (maybe with task to study it before the course) and more practice would have been more useful to me (that I was basically ignorant of these topics). Congratulations to the organizers. This initiative is very precious for the INFN community. Complimenti e grazie mille.

Conclusion

Overall, the second hackathon was successful.

- All applicants were able to participate
- Broad spectrum of participants from the INFN Community
- Positive feedback: level was appropriate
- The technical setup worked flawlessly.

Opportunities to make it better next time

- Most requests are for making it longer. Adding an empty day before the hackathon?
- Trying to balance the level of participants in a group, rather than the level of groups
- Preparing guidelines for self “post-training” with pointers for additional studies