Nuclear Talent Course on Machine Learning and Data Analysis for Nuclear Physics, ECT*, June 22-July 3, 2020

I. INTRODUCTION TO THE TALENT COURSES

A recently established initiative, Training in Advanced Low Energy Nuclear Theory, aims at providing an advanced and comprehensive training to graduate students and young researchers in low-energy nuclear theory. The initiative is a multinational network between several European and Northern American institutions and aims at developing a broad curriculum that will provide the platform for a cutting-edge theory for understanding nuclei and nuclear reactions. These objectives will be met by offering series of lectures, commissioned from experienced teachers in nuclear theory. The educational material generated under this program will be collected in the form of WEB-based courses, textbooks, and a variety of modern educational resources. No such all-encompassing material is available at present; its development will allow dispersed university groups to profit from the best expertise available.

The present document aims at giving a summary, with background information as well, of the Nuclear Talent course, *Machine Learning and Data Analysis for Nuclear Physics*. The course was held online via the premises of the European Center for Theoretical Studies in Nuclear Physics and Related Areas (ECT*), Trento, Italy during the period June 22 to July 3, 2020.

This was the sixth course which has been held at the ECT*, the first one in 2012, then 2014, 2015, 2017, 2019 and finally this one in 2020. Although we had to run this course as a fully online course due to the global COVID-19 pandemic, the support from the ECT* has simply been crucial for the successful arrangement and running of these courses. All practical arrangements, from accepting applications to the practical set up of the online course, were simply central for the success of the present course held in the summer of 2020. In total we had 157 participants from countries in Africa, Asia, Europe. Oceania and the Americas. The daily attendance varied from 110 to close to 150.

II. REPORT

A. Introduction

Probability theory and statistical methods play a central role in science. Nowadays we are surrounded by huge amounts of data. For example, there are about one trillion web pages; more than one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day; the genomes of 1000s of people, each of which has a length of more than a billion base pairs, have been sequenced by various labs and so on. This deluge of data calls for automated methods of data analysis, which is exactly what machine learning provides. The purpose of this Nuclear Talent course was to provide an introduction to the core concepts and tools of machine learning in a manner easily understood and intuitive to physicists and nuclear physicists in particular. We started with some of the basic methods from supervised learning and statistical data analysis, such as various regression methods before we moved into deep learning methods for both supervised and unsupervised learning, with an emphasis on the analysis of nuclear physics experiments and theoretical nuclear physics. The students worked on hands-on daily examples as well as projects that resulted in final credits. The major scope was to give the participants a deeper understanding on what Machine learning and Data Analysis are and how they can be used to analyze data from nuclear physics experiments and perform theoretical calculations of nuclear many-body systems.

B. Aims and outcomes

The two-week online TALENT course on nuclear theory focused on Machine Learning and Data Analysis algorithms for nuclear physics and how to use such methods in the interpretation of data on the structure of nuclear systems.

We proposed approximately twenty hours of lectures over two weeks and a comparable amount of practical computer and exercise sessions, including the setting of individual problems and the organization of various individual projects.

The mornings consisted of lectures and the afternoons were devoted to exercises meant to shed light on the exposed theory, the computational projects and individual student projects. These components will be coordinated to foster student engagement, maximize learning and create lasting value for the students. For the benefit of the TALENT series and of the community, material (courses, slides, problems and solutions, reports on students' projects) have been made publicly available using version control software like git and posted electronically on github (this site).

At the end of the course the students should have obtained a basic understanding of

- Statistical data analysis, theory and tools to handle large data sets.
- A solid understanding of central machine learning algorithms for supervised and unsupervised learning, involving linear and logistic regression, support vector machines, decision trees and random forests, neural networks and deep learning (convolutional neural networks, recursive neural networks etc)
- Be able to write codes for linear regression, logistic regression and use modern libraries like Tensorflow, Pytorch, Scikit-Learn in order to analyze data from nuclear physics experiments and perform theoretical calculations
- A deeper understanding of the statistical properties of the various methods, from the bias-variance tradeoff to resampling techniques.

We targeted an audience of graduate students (both Master of Science and PhD) as well as post-doctoral researchers in nuclear experiment and theory and due to the online format we had in total 157 participants.

The teaching team consisted of both theorists and experimentalists. We believe such a mix is important as it gives the participants a better understanding on how data are obtained, and what are the limitations and possibilities in understanding and interpreting the experimental information.

C. Course Schedule

Lectures were approximately 45 min each with a small break between each lecture. The schedule is included here, with links to online material as well.

1. Week 1

- Monday June 22: Linear Regression and intro to statistical data analysis (Morten Hjorth-Jensen MHJ). Learning slides at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Introduction/html/Introduction.html` and `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day1/html/Day1.html` and link to video from lecture June 22 `https://mediaspace.msu.edu/media/t/1_ogq38oqq`
- Tuesday June 23: Logistic Regression and classification problems, intro to gradient methods (MHJ). Learning slides at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day2/html/Day2.html` and link to video for first lecture at `https://mediaspace.msu.edu/media/t/1_po1a5e9v` and second lecture at `https://mediaspace.msu.edu/media/t/1_wbz4v2gm`
- Wednesday June 24: Decision Trees, Random Forests and Boosting methods (MHJ). Learning slides at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day3/html/Day3.html` and link to video at `https://mediaspace.msu.edu/media/t/1_vrt5rxls`
- Thursday June 25: Basics of Neural Networks and writing your own Neural Network code (MHJ). Learning slides at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day4/html/Day4.html` and link to video at `https://mediaspace.msu.edu/media/t/1_ksuz0ero`. The link to the video of the additional exercise session is at `https://mediaspace.msu.edu/media/t/1_shte4iw5`
- Friday June 26: Beta-decay experiments, how to analyze various events, with hands-on examples . (Sean Liddick) Videos and teaching material `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day5/html/Day5.html`. Link to video of online lecture at `https://mediaspace.msu.edu/media/1_5n2bssbl`. The link to the video of the additional exercise session is at `https://mediaspace.msu.edu/media/1_q74f31cw`

2. Week 2

- Monday June 29: Neural Networks and Deep Learning (Raghu Ramanujan, RR). PDF file of the presented slides at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day6/pdf/Day6.pdf`. Jupter-Notebook at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day6/ipynb/Day6.ipynb`. Video of lecture at `https://mediaspace.msu.edu/media/t/1_58a9xrbt`. Video of exercise session at `https://mediaspace.msu.edu/media/t/1_ulont3rg`

- Tuesday June 30: From Neural Networks to Convolutional Neural Networks and how to analyze experiment (classification of events and real data) (Michelle Kuchera, MK). Jupyter-notebook of lecture at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day7/ipynb/Day7.ipynb`. Video of lecture `https://mediaspace.msu.edu/media/t/1_2ysd5plh` and video of exercise at `https://mediaspace.msu.edu/media/t/1_watjxppf`

- Wednesday July 1: Discussion of nuclear experiments and how to analyze data, presentation of simulated data from Active-Target Time-Projection Chamber (AT-TPC) (Daniel Bazin). Slides of lectures (PDF) at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day8/pdf/Day8.pdf`. Videos and teaching material `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day8/html/Day8.html`. Video of actual lecture at `https://mediaspace.msu.edu/media/t/1_azaquoc0`. Video of analysis of data with CNNs (MK) at `https://mediaspace.msu.edu/media/t/1_rozywc7h`. Jupyter-notebook of hands-on session at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day8/ipynb/Day8.ipynb`

- Thursday July 2: Generative models (MK). Slides of lectures (PDF) at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day9/pdf/Day9.pdf`. Jupyter-notebook at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day9/ipynb/Day9.ipynb`. Video of lecture at `https://mediaspace.msu.edu/media/t/1_ayfst99b`. Video of exercise session at `https://mediaspace.msu.edu/media/t/1_wpdmt7cw`.

- Friday July 3: Reinforcement Learning (RR). Slides of lectures (PDF) at `https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day10/pdf/Day10.pdf`. Future directions in machine learning and summary of course. Video of first lecture at `https://mediaspace.msu.edu/media/t/1_0eiikln6`. Video of second lecture at `https://mediaspace.msu.edu/media/t/1_wzyabacr`.

D. Teaching

The course was taught as an online intensive course of duration of two weeks, with a total time of 20 h of lectures and 10 h of exercises, questions and answers. Videos and digital learning material were made available one week before the course begins. It was possible to work on a final assignment of 2 weeks of work. The total load wasapproximately 80 hours, corresponding to 5 ECTS in Europe. The final assignment were graded with marks A, B, C, D, E and failed for Master students and passed/not passed for PhD students. A course certificate was issued for students requiring it from the University of Trento.

The organization of a typical course day was as follows:

1. *Time and Activity*

- 2pm-4pm (Central European time=CET) Lectures, project relevant information and directed exercises

- 5pm-6pm (CET) Questions and answers, Computational projects, exercises and hands-on sessions

E. Teachers and organizers

The teachers and organizers were

1. Daniel Bazin at Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan, USA (DB)

2. Morten Hjorth-Jensen at Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan, USA (MHJ)

3. Michelle Kuchera at Physics Department, Davidson College, Davidson, North Carolina, USA (MK)

4. Sean Liddick at Department of Chemistry and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan, USA (SL)

5. Raghuram Ramanujan at Department of Mathematics and Computer Science, Davidson College, Davidson, North Carolina, USA (RR)

Morten Hjorth-Jensen functioned as student advisor and coordinator.

III. BRIEF SUMMARY

The teachers were fully self-sponsored, and due to the COVID-19 situation worldwide there were no travel or logding expenses. The course was fully online.

The support from the ECT*, and its year-long experiences with running doctoral training programs and Talent courses, was central to the success of the course. Of uttermost importance was Barbara Gazzoli's help will all administrative matters. Without her help and the other staff of the ECT*, and the director's (Prof. Jochen Wambach) enthusiastic support, it is unlikely that this course would have run as smoothly as it did, or at all!

Overall, we as teaching team feel the original learning outcomes and goals were met. It was a marvelous group of students to teach. Videos of the lectures will be published by Springer in due time.

This time we did however not perform a course survey, feedback from students is thus missing.