Machine Learning In Astronomy

Dr. Michelle Lochner

1 Introduction

Welcome to a crash course on machine learning in astronomy and cosmology! We will cover some of the basics of machine learning, just enough to give you an idea of how to go about solving problems with basic machine learning tools. As with all my courses, this course focuses on self-directed learning but we will have extensive in class activities with a final report and presentation due at the end of the course. Classes are on a Monday at 10am until 12pm in the board room.

The marks for this course are as follows:

- Quiz on video materials (29 September) - 10%
- Supernova classification project report (due 10 October)- 60%
- Presentations on machine learning algorithms (13 October) - 30%

2 Schedule

2.1 22 September - NO LECTURE

Use this time instead to work through the following set of videos. These are pre-recorded lectures I gave at the European Space Agency. Although they are old by now, the fundamentals of machine learning have not changed and must be mastered before moving on to more advanced topics. There are approximately 3 hours of lectures split over 3 videos. The deep learning lecture (final video) is optional, but highly recommended if you are interested in taking the topic further.

1. https://www.youtube.com/watch?v=ubhLmFQPgjc&t=4907s
2. https://www.youtube.com/watch?v=le4w_YY4TxQ
3. https://www.youtube.com/watch?v=vRTbRwqDqUA&t=1839s

The slides are also available at this link:

https://github.com/MichelleLochner/ml-tutorials/blob/main/lecture_notes.pdf.

Be sure to familiarise yourself with the material *before* arriving at the next class.

2.2 29 September - In-class Quiz & Tutorial

We will start the session with a short quiz on the videos you watched. Then we will spend time working on machine learning tutorials. In preparation for this, download the files from this github repository to your computer: https://github.com/MichelleLochner/ml-tutorials. Click the green "code" button and then click "download zip" to download the zip file to your computer where you can unzip it.

On the same link, if you scroll down you will see some instructions on required packages to be able to run the tutorials. You can install these packages by typing this on the command line:

```
pip install --user scipy astropy sncosmo jupyter ipympl scikit-learn iminuit
```

Make sure you can launch the `tutorial-basics.ipynb` tutorial before coming to class. In class we will work through this tutorial and you will have the opportunity to ask any questions. If you finish it, you can start on `tutorial-deep-learning.ipynb`. By the end of class, you should have a good idea of how to solve a basic classification problem with machine learning. You should also ensure you can run the `tutorial-supernovae.ipynb` notebook before leaving class. Your main project will be based on this notebook and it's a good idea to start working on it as soon as possible.

2.3 6 October - Supernova Classification Project

In a separate file, I have written a description of the project that will form the bulk of your marks for this course. You should start your project as soon as you understand the lectures and the basic tutorial, *don't wait until the 6th of October to start working on it!*.

The lecture slot on the 6th of October is reserved for helping you with your project, the due date of which is the 10th of October at 10pm. **Late hand in is only possible until the 12th of October 10pm.** This is because your reports must be in before the presentation. Handing in after the deadline but before the 11th of October 10pm will incur a mark cap of 80% for your report, handing in before the 12th of October 10pm will incur a mark cap of 60%. No further late hand-ins will be possible.

2.4 13 October - Presentations

For your projects, each of you has been assigned a different classification algorithm (which you can find on the instructions pdf). On the 13th of October you will make a presentation describing your algorithm. You will also present any analysis choices you made, such as the size of your training set and choice of hyperparameters, and your final results. Your presentation must be 12 minutes, with 3 minutes for questions from myself and the class. You will be timed and I will cut off students who go over time. You will be marked on the following (marks given in brackets):

- *(10)* Clarity of algorithm explanation (demonstrating understanding of the algorithm)
- *(6)* Clarity of plots demonstrating results (clearly readable plots that make sense)
- *(6)* Overall quality of presentation slides (uncluttered slides, avoiding lots of bullet points, clean formatting etc.)
- *(4)* Overall clarity of oral presentation (not speaking too fast or too softly, eye contact with audience, conveying points clearly and concisely, without loss of information)
- *(4)* Keeping to time (not going over time but also not too many minutes under time)

Be sure to practice your talk before attending class!