International Workshop
on Machine Learning
for Space Weather:
Fundamentals, Tools and
Future Prospects

7-11 November 2022
This is a hybrid meeting
Buenos Aires, Argentina

Further information:
https://indico.ictp.it/event/9640/
smr3750@ictp.it
+39-040-2240284
Elizabeth Brancaccio

The Abdus Salam
International Centre
for Theoretical Physics
www.ictp.it
Trieste, Italy

Hands on #1
Classification (using Keras)

Problem #1:

There are different methods of monitoring the ionosphere, one of the most used is with radar systems that take advantage of the reflective characteristics of the ionosphere and the properties of the propagation of electromagnetic waves. A problem that can be applied to different radar systems has to do with the ability to classify the echoes received at the radar antenna. In this exercise it is proposed to analyze a set of data corresponding to received echoes. When they are the result of a reflection in the ionosphere, they should be classified as a valid echo. The "ionosphere.data" file is available with recorded echoes that are already classified (positive return or not) manually by experts.

For this problem:
1. You need to build a classifier that allows to determine, based on the echoes (features) available from the radar, whether it is a positive echo or not (target).
2. During the data engineering stage:
    a) Determine if the data is balanced.
    b) Propose some technique for its balancing in case there is a significant imbalance.
3. Analyze the results from several generated models.
4. Show plots of loss function and accuracy of each of the models generated.

Problem #2:

Coronal Mass Ejections (CMEs) are one of the most important events that occur on the Sun. In many cases, the origin of CMEs can be the so-called solar flares or "flares", although there are other events that give rise to origin to CMEs (e.g. filaments). The file "CMEs.csv" contains a list of CMEs cataloged according to their origin (flare/no-flare), and also has two features corresponding to the solar activity at the time of occurrence of the CME (low/high) and the propagation speed. The columns that the file has are: "SunActivity", "vprop" and "Origin".
    • "SunActivity": categorical variable that indicates solar activity in the year the record was taken (low, high).
    • "Vprop": numerical variable that indicates the speed of the solar wind during the recording.
    • "Origin": binary variable that indicates whether or not there was a flare (flare, no flare) before a CME.

For this problem:
1. You need to build a classifier on the "Origin" variable.
2. Determine if the data is balanced.
3. What kind of categorization of variables is necessary?

International Workshop on Machine Learning for Space Weather: Fundamentals, Tools and Future Prospects

7-11 November 2022
This is a hybrid meeting
Buenos Aires, Argentina

Further information:
https://indico.ictp.it/event/9640/
smr3750@ictp.it
+39-040-2240284
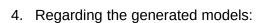Elizabeth Brancaccio

4. Regarding the generated models:
   a) What issues can be identified in the data that hurt training?
   b) Do they have an acceptable precision of the models according to your point of view?
   c) What alternatives would you propose to improve the models?
5. Analyze the results from several generated models.
6. Show plots of loss function and accuracy of each of the models generated.