

# ADVANCED MACHINE LEARNING

## PROJECT PROPOSAL: Magic Telescope

Omar López Rubio  
Maria Paraskeva

The dataset we propose for the course project of AML can be found here: [Datasets](#) with the title “Magic Telescope”.

It is a classification problem that deals with Gamma particles. The reason behind our choice is that we think it is an interesting data set that will allow us to implement the contents of this course. After reading some proposals of solutions for the classification task found in Kaggle, we want to see if we can accomplish better results or implement different methods for solving it.

In particular, we are interested in using SVM with custom kernels and comparing it to Naive Bayesian.

The data are MC generated to simulate registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. Cherenkov gamma telescope observes high energy gamma rays, taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas, and developing in the atmosphere. This Cherenkov radiation (of visible to UV wavelengths) leaks through the atmosphere and gets recorded in the detector, allowing reconstruction of the shower parameters. The available information consists of pulses left by the incoming Cherenkov photons on the photomultiplier tubes, arranged in a plane, the camera. Depending on the energy of the primary gamma, a total of few hundreds to some 10000 Cherenkov photons get collected, in patterns (called the shower image), allowing to discriminate statistically those caused by primary gammas (signal) from the images of hadronic showers initiated by cosmic rays in the upper atmosphere (background).

Characteristics of the data set:

- ❖ 12 features (11 numerical and 1 categorical with 2 classes)
- ❖ 19020 total instances
- ❖ 0 missing values

References:

[arXiv:1909.01172](#) [astro-ph.IM]