Predecir Clasificación de Rayos de Gama del MAGIC Gamma

María José Soto Castro

A01705840@tec.mx

1. INTRODUCTION

To understand how telescopes like MAGIC work, it's essential to first grasp the nature of the high-energy particles and radiation they are designed to detect.

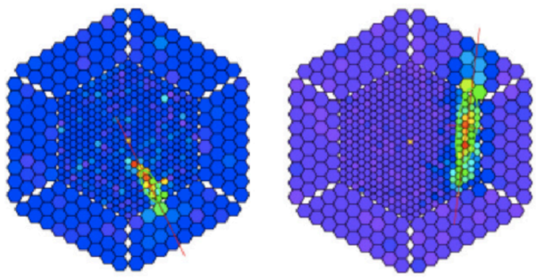
- **Gamma Rays:** Gamma rays are the most energetic form of electromagnetic radiation, with individual photons having energies above 100 keV. Unlike visible light, they are completely absorbed by the Earth's atmosphere. This makes direct observation from the ground impossible. However, their interaction with the atmosphere is precisely what allows ground-based telescopes to detect them indirectly.
- **Hadron Rays:** Hadrons are a class of composite particles, which include protons and neutrons. Hadron rays, or more commonly, "hadronic cosmic rays," are high-energy hadrons that continuously bombard the Earth's atmosphere. Like gamma rays, they are also absorbed by the atmosphere, initiating a cascade of secondary particles

When a high-energy gamma ray or a cosmic-ray hadron enters the atmosphere, it triggers a chain reaction known as an "extensive air shower." This shower is a cascade of secondary particles, and it's what ground-based telescopes actually observe. The challenge is that both gamma rays and hadrons produce these showers, and they can be difficult to distinguish.

2. DATASET DESCRIPTION

It is a collection of 19,020 simulated instances representing these showers, each described by 10 features known as Hillas parameters. These parameters mathematically characterize the shape and orientation of the shower's image, allowing a classification algorithm to learn the subtle differences between the regular, elliptical showers from gamma rays and the more irregular, dispersed showers from hadrons.

An interpretation of the data on the camera would be something as the next image:



Camera images produced by different EAS. Left: γ-ray initiated shower (gamma). The compact ellipse points to the source direction. Right: Hadronic shower, broader than the electromagnetic shower ellipse and with an arbitrary direction. (López-Oramas, Alicia., 2015)

3. DATA CLEANING

First, the code loads the raw data from a file named magic04.data and assigns meaningful column names. Next, it shuffles the entire dataset randomly to ensure that the distribution of classes is uniform across all data splits. The core of the cleaning involves feature scaling: a custom normalize function is applied to all feature columns (excluding the class label) to scale their values to a range between 0 and 1. Using this formula: This normalization is crucial for algorithms like gradient descent, which can perform poorly with features on different scales. Finally, the class labels,

Predecir Clasificación de Rayos de Gama del MAGIC Gamma

María José Soto Castro

A01705840@tec.mx

originally 'g' and 'h' (representing gamma and hadron events), are converted into a binary numerical format, with 'g' mapped to 0 and 'h' mapped to 1. This pre-processing makes the data suitable for the logistic regression model that follows.

4. MODEL USED

The machine learning model employed for the classification of gamma and hadron events is a custom implementation of **Logistic Regression** trained using **Batch Gradient Descent**. Logistic Regression is a statistical model that, in its most basic form, uses a logistic function to model a binary dependent variable. It estimates the probability of a given instance belonging to a particular class (in this case, either a gamma ray or a hadron).

5. TRAINING, VALIDATION AND TESTING

The dataset, containing 19,020 instances, was split into three distinct subsets to ensure a robust evaluation of the model's performance:

Training Set (60%): During each epoch, the model's parameters were updated by applying Batch Gradient Descent to minimize the binary cross-entropy loss function.

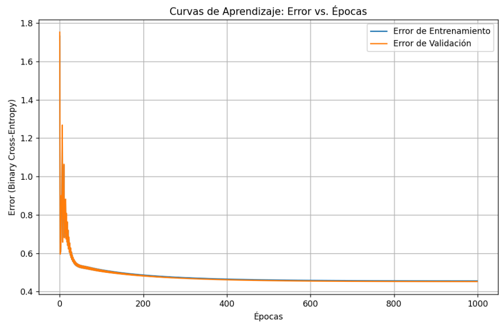
Validation Set (20%): The validation error was calculated at each epoch to check for overfitting and to assess how well the model generalizes to unseen data. This set helps determine when to stop the training process to prevent the model from becoming too specialized to the training data.

Test Set (20%): This set was not used during training or validation and provides an unbiased measure of the model's final performance, specifically its accuracy on truly new, unseen data.

The model was trained for up to 500 epochs with a learning rate of 0.005, and the training was set to stop if the training error dropped below a predefined threshold of 0.01.

6. RESULTS

Based on the analysis of the learning curves and prediction histograms, the model is currently underperforming due to a high bias. Both the training and validation error curves converge and stabilize at a relatively high value of around 0.45. The minimal distance between these two curves indicates that the model is not memorizing the training data, resulting in a low variance.

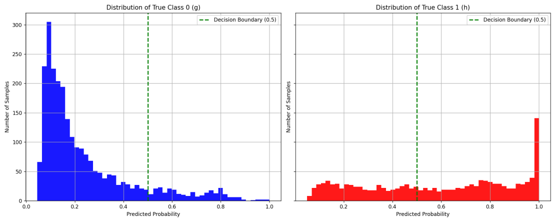


The model's high bias diagnosis is further supported by the prediction histograms. Ideally, the distributions for the two classes should be well-separated. However, the graphs show significant overlap, particularly in the center around the decision boundary. This indicates the model lacks capacity to effectively separate the classes, leading to numerous incorrect predictions and a high overall error.

Predecir Clasificación de Rayos de Gama del MAGIC Gamma

María José Soto Castro

A01705840@tec.mx



REFERENCES

Conceição, R., Costa, P. J., Gibilisco, L., Pimenta, M., & Tomé, B. (2023). : a high resolution gamma/hadron and composition discriminant variable for Water-Cherenkov Detector cosmic-ray observatories

MAGIC – INAF Osservatorio Astronomico di Brera. (2019). Inaf.it. https://brera.inaf.it/en/progetti_ricerca/magic/#:~:text=SHORT%20DESCRIPTION,physics%20and%20high%2Denergy%20astrophysics.

Probing the Cosmic Ray Background of Gamma-Ray Astronomy with Hadron Colliders. (2017). Arxiv.org. https://arxiv.org/html/2509.04040v1#:~:text=These%20CR%20particles%2C%20composed%20predominantly,form%20of%20hadronic%20air%20showers.

López-Oramas, Alicia. (2015). Multi-year Campaign of the Gamma-Ray Binary LS I +61◦ 303 and Search for VHE Emission from Gamma-Ray Binary Candidates with the MAGIC Telescopes. 10.13140/RG.2.1.4140.4969.