ITESO, Universidad Jesuita de Guadalajara

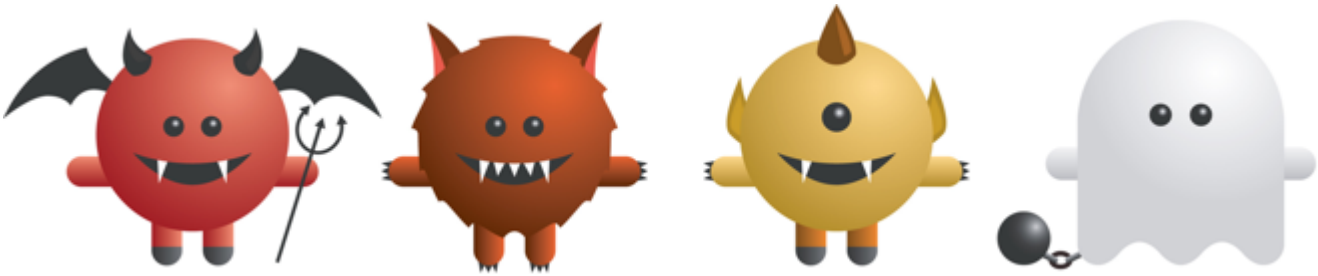
March Saturday 20, 2021

Machine Learning Exam 1

Ghouls, Goblins, and Ghosts... Boo!

Description

Get out your dowsing rods, electromagnetic sensors, … and gradient boosting machines. Kaggle is haunted and we need your help. After a month of making scientific observations and taking careful measurements, we've determined that 900 ghouls, ghosts, and goblins are infesting our halls and frightening our data scientists. When trying garlic, asking politely, and using reverse psychology didn't work, it became clear that machine learning is the only answer to banishing our unwanted guests.



So now the hour has come to put the data we've collected in your hands. We've managed to identify 371 of the ghastly creatures, but need your help to vanquish the rest. And only an accurate classification algorithm can thwart them. Use bone length measurements, severity of rot, extent of soullessness, and other characteristics to distinguish (and extinguish) the intruders. Are you ghost-busters up for the challenge?

Evaluation

Submissions are evaluated on the categorization accuracy (the percent of creatures that you correctly classify).

Submission File

Your submission file should predict the type for each creature in the test set. The file should contain a header and have the following format:

```
id,type
0,Ghost
1,Goblin
2,Ghoul
etc.
```

ITESO, Universidad Jesuita de Guadalajara

March Saturday 20, 2021

Instructions

Create an account in www.kaggle.com, use the **username** same as **Iteso email**.

Go to next link to has more details and dataset access.

https://www.kaggle.com/c/ghouls-goblins-and-ghosts-boo

Create a model (one vs all or softmax) considering all the topics seen in class. Once you have a solution, submit the file in the appropriate format and get your final score.

Canvas

Upload to canvas the files including your notebook (html, ipynb) document describing all steps, modifications to dataset, how optimization was done, conclusion, etc.

The notebook must include a screenshot where your score is shown from submission as follows.

Notes:

- Your code should create a file with your predictions of test dataset submitted to kaggle.
- Do not upload only your source code without the **html** document.
- You can do more than one submission, take your highest score.
- I will run your notebook and submit the created file to verify the score.
- You can add or remove features