Name: Christian Hellum Bye   Github: christianhbye   TPM: Rick Wu

The project is the pre-approved "Predicting a pulsar star".

Dataset

The data is from https://www.kaggle.com/pavanraj159/predicting-a-pulsar-star [1]. I chose this dataset because I have a strong interest in astrophysics and doing a project like this is similar to some astrophysics research. In fact, the research I already do is strongly concerned with radio frequency interference (RFI), so it would be interesting to see how RFI is dealt with in different experiments within the same field.

Methodology

The data chosen is already suited for training a machine learning model. It solely consists of numbers and has only eight parameters – i.e. it is not crucial to minimize/remove parameters to optimize the model.

Based on eight parameters, we want to predict whether a signal is RFI or a real pulsar star. Essentially, we have nine coloumns and we want to be able to predict the entries of the ninth – is it a 0 (RFI) or 1 (real star)?

For the machine learning model, a classifier algorithm (maybe from this list) seems appropriate since this is a discrete problem. However, at the current time we do not know enough about the different models to have a good justification to pick one over the other, but hopefully will change after some more lectures. Logistic regression seems like one potential candidate though.

Final conceptualization

I will make a poster to present my results. The data contain a total of 17,898 possible pulsar stars of which 1,639 are positive examples (a pulsar star) and 16,259 negative examples (not a pulsar star). Thus, a model that predicts no pulsar stars will achieve about 91 % accuracy, so we should at least try to beat that by a margin. Other contributors (such as Pavan Raj [2] on Kaggle) predicted 9 % pulsar stars which is at most off the correct value by less than one percentage point, whereas the user Efe Ergun [3] on Kaggle achieved more than 97 % accuracy with a logistic regression.

References

[1] Raj, Pavan. (2018). Predicting a Pulsar Star. [online] Kaggle.com. Available at: https://www.kaggle.com/pavanraj159/predicting-a-pulsar-star [Accessed 28 Jan. 2020].

[2] Raj, Pavan. (2018). predicting a pulsar star in the universe. [online] Kaggle.com. Available at: https://www.kaggle.com/pavanraj159/predicting-a-pulsar-star-in-the-universe [Accessed 28 Jan. 2020].

[3] Ergun, Efe. (2019). Predicting A Pulsar Star. [online] Kaggle.com. Available at: https://www.kaggle.com/efeergun96/prediciting-a-pulsar-star [Accessed 28 Jan. 2020].