

# Pulsar detection - Machine learning and pattern recognition exam

Alberto Baroso - s296520

2022-07-13

## Abstract

The Pulsar dataset [1] is a collection of 17,898 pulsar candidates of which 1,639 are real pulsars and 16,259 are spurious examples caused by RFI/noise. A pulsar is a neutron star that emits beams of electromagnetic radiation. As pulsars rotate, their emission beam sweeps across the sky, a periodically repeated pattern of broadband radio emission can be detected as this beam crosses our line of sight. Thus pulsar search involves looking for periodic radio signals with large radio telescopes. The goal of this project is to use machine learning techniques to solve the binary classification problem of detecting pulsars from the Pulsar dataset.

## Contents

<b>1</b>	<b>Problem analysis</b>	<b>1</b>
1.1	Features . . . . .	1
1.2	Feature distributions and ranges . . . . .	2

## 1 Problem analysis

### 1.1 Features

Pulsar candidates are described by 8 continuous features, and a class label.

The first four are simple statistics obtained from the integrated pulse profile (folded profile). This is an array of continuous features that describe a longitude-resolved version of the signal that has been averaged in both time and frequency (see [2] for more details). The remaining four features are

similarly obtained from the DM-SNR curve (again see [2] for more details). These are summarised below:

1. Mean of the integrated profile.
2. Standard deviation of the integrated profile.
3. Excess kurtosis of the integrated profile.
4. Skewness of the integrated profile.
5. Mean of the DM-SNR curve.
6. Standard deviation of the DM-SNR curve.
7. Excess kurtosis of the DM-SNR curve.
8. Skewness of the DM-SNR curve.
9. Class

The class labels used are 0 (negative) and 1 (positive).

## 1.2 Feature distributions and ranges

There are 8 features, that represent statistics extracted from radio signals collected by radio telescopes. Classes are highly imbalanced.

## References

- [1] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, and J. D. Knowles. Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, 459(1):1104–1123, 04 2016.
- [2] R. J. Lyon. *Why are pulsars hard to find*. PhD thesis, University of Manchester, 2016.