

UNIVERSITY OF ST ANDREWS

MACHINE LEARNING

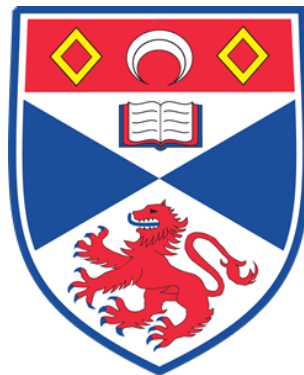
CS5014

Classification

Author:

150008022

April 13, 2019



Goal

The goal of this practical is to analyse a dataset in order to produce a classification model that can make predictions based on a set of inputs.

Contents

1	Loading Data	1
2	Cleaning Data	1
3	Data Visualisation and Analysis	1
3.1	Distributions	2
3.2	Relationships	2
4	Feature Selection	2
5	Model Selection and Training	2
5.1	Model 1	2
5.2	Model 2	2
6	Evaluation and Comparison	2
7	Discussion	2

1 Loading Data

To load the data, the paths to the relevant files are supplied as arguments to the `__main__.py` script. The *pandas* module was used to load the file contents into *DataFrames*.

A test set was isolated from the original data using an 80%-20% split. Stratification was used to ensure that all classes were represented in the training data.

2 Cleaning Data

When originally loading the CSV files the parameter to raise an exception on missing or extra columns was included, and so it could be assumed that all rows had the same number of columns. The `dtype=float` argument was also passed when loading the data to ensure that each column contained the expected numerical data. Any rows containing empty or NaN values were dropped from the dataset.

3 Data Visualisation and Analysis

The input CSV was understood to have the structure shown in figure 1 repeated once for each of the four channels.

Index	Channel 1												
	Component 1			Component 2			Component 4			...	Component 64		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max		Mean	Min	Max
1													

Figure 1: The structure of each row of the CSV file which is repeated for each channel.

3.1 Distributions

3.2 Relationships

4 Feature Selection

The practical specification mentions that the data was collected in a slightly different way to the RadarCat paper [1].

Similar features

5 Model Selection and Training

5.1 Model 1

5.2 Model 2

6 Evaluation and Comparison

7 Discussion

References

- [1] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. Radarcats: Radar categorization for input & interaction. pages 833–841, 10 2016.