CS5014 Machine Learning

Classification of object colour using optical spectroscopy

April 19, 2018

Lecturer:
David Harris-Birtill
Kasim Terzić

Submitted By:
140011146

1   Introduction

In this practical, two classification tasks were performed on an experimental dataset.

2   Binary Classification

2.1   Preprocessing

Before any preprocessing is done, the data was split into training and test datasets. This is to ensure no bias when looking at the data, even for visualisations.

2.1.1   Data visualisation

First, the spectroscopy data can be visualised by plotting the intensities and wavelengths to see if there are visual patterns than can be seen in the data. As there are over nine hundred input features, it is likely many of them are either redundant or not well correlated to the input data.
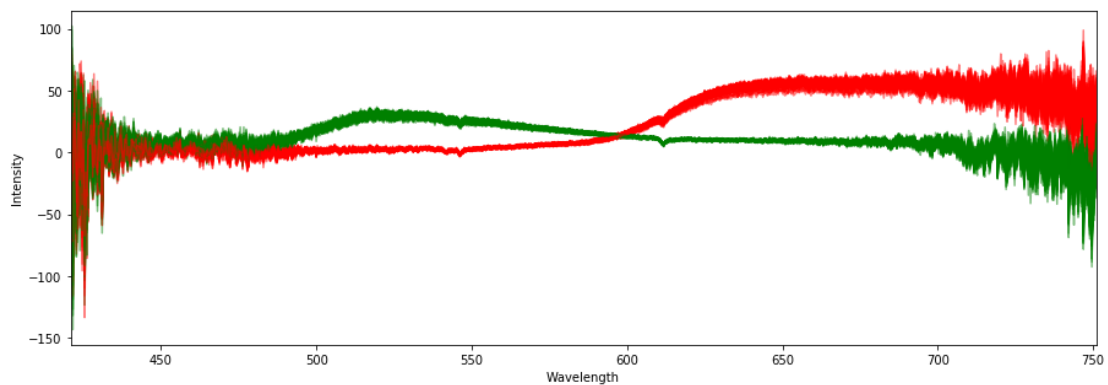

Figure 1: Initial plot of the data, showing the intensity of the wavelengths measured. The plot is colour coded with the output so any differences between the red and green data can be easily seen.

From the visualisation, it is interesting to see that for many wavelengths, there is a very clear distinction between the red and green intensities. This suggests that for those wavelengths, it would be very simple to classify. Further, the visualisation was created from all the training data, so it can be seen there are no large errors in the middle wavelengths where no data points incurred significant differences, which shows the consistency of the data. It can be see towards the start and end of the spectrum, the intensities are more varied, indicating possible errors during measurement. As such, it would make sense when choosing input features to mostly include features from the middle wavelengths where less possible errors occurred.

2.1.2   Data correlation

Next, the correlation of each input feature wavelength can be calculated and plotted to see if the set of input features in the middle are well correlated to the output.
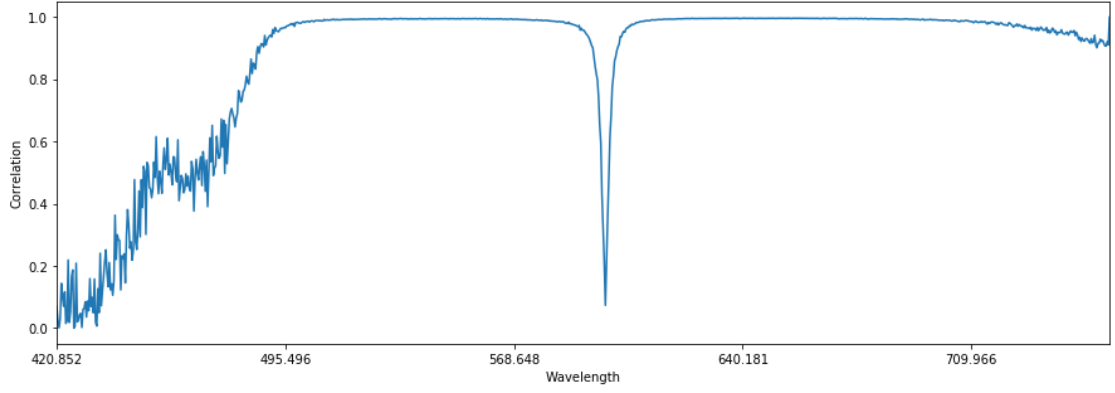
Figure 2: Correlation of each wavelength input feature to the output classification.

The curve of input feature correlation corresponds very closely to figure 1. At the wavelengths where the red and green curves become distinct from each other, the correlation of that wavelength to the output class tends to perfect correlation. At the center where the red and green curves cross over, a dip in the correlation can be seen as it becomes difficult to distinguish the two in the overlap.
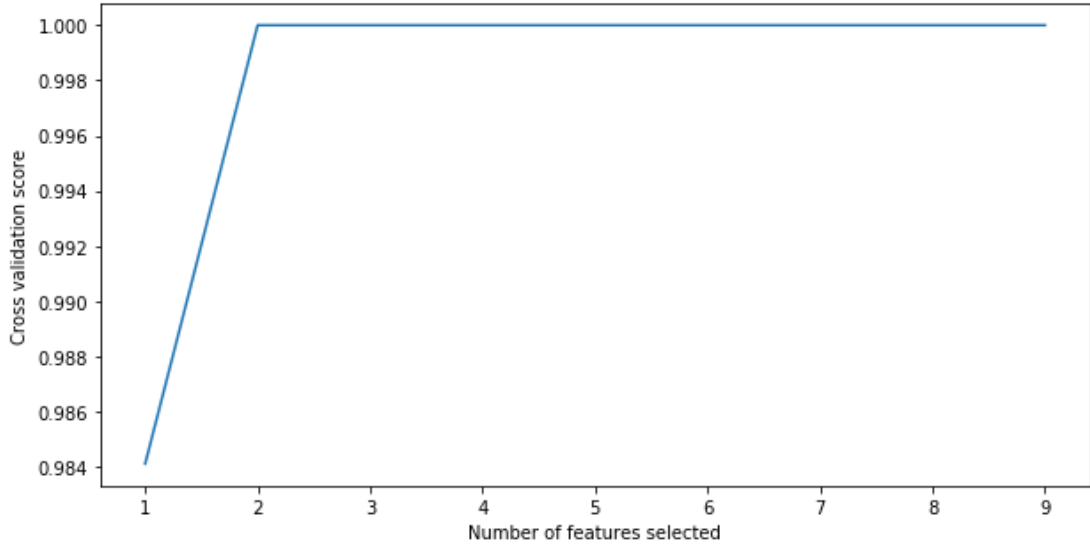
2.2 Results

Figure 3: Cross validation accuracy score of increasing number of features using recursive feature elimination.

3   Multiclass Classification

3.1   Preprocessing

3.1.1   Data visualisation

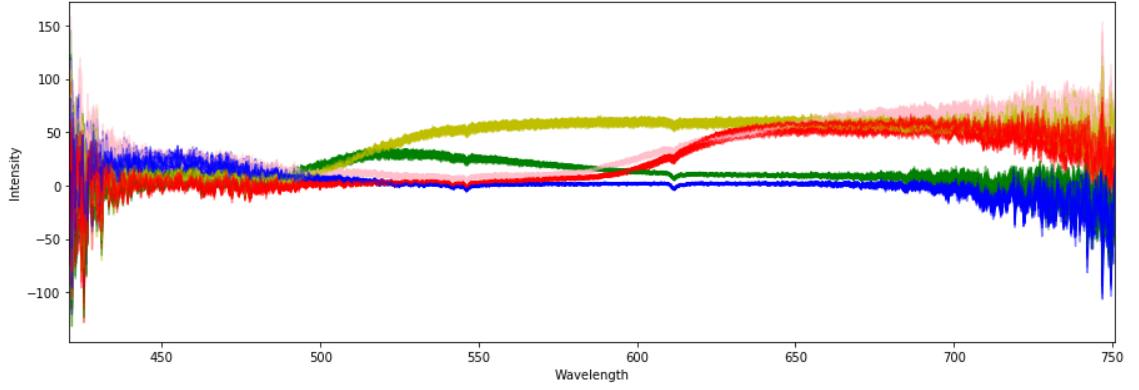Figure 4: Plot of the data to show the intensity of wavelengths of the different colour classes.
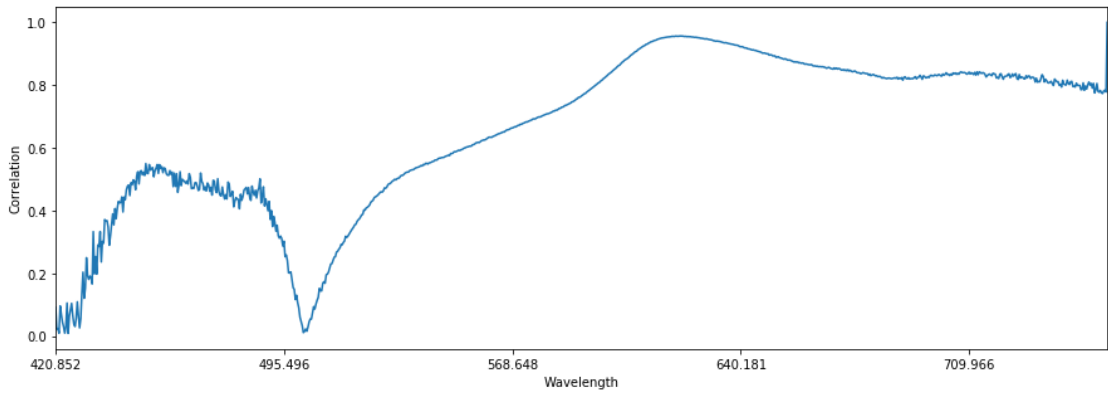
3.1.2   Data correlation

Figure 5: Correlation of each wavelength input feature to the colour classification.

3.2   Results

4   Conclusion and Evaluation