CS5014 - Machine Learning

P2 - Classification of object colour using optical spectroscopy

Report

140014952

April 19, 2018

1 Overview

The objectives of this practical were to come up with classification model for binary and multi-class classification problems. This submission investigates both binary and multi-class tasks. The solution python scripts can be found in */binaryML/* and */multiclassML/* directories. Corresponding predicted class files are also in these directories.

As later sections suggest, the amount of features it takes to determine the class for each task is very low. This is hypothesized based on input analysis and later machine learning observations support these hypotheses.

2 Cleaning Data and Feature Extraction

As the very first step input data were split into training and testing data with 70-to-30 ratio. The analysis were first done on the training set.

(negative values)

3 Data Analysis and Visualization

To visualize training data X it was plotted with provided wavelength data that contains information about each feature wave length. Additionally, Y training set was used to indicate visually how colours are distributed and can be distinguished. This gave insights as to which features are likely to be good indicators for predicting the colour. Plots for both binary and multi-class data were generated.

3.1 Binary Classification

Fig.1 shows that there is a clear distinction between red and green colours in terms of features. From the same figure it can be said feature values between 400 and 500 wavelength are more or less shared between both red and green colours. Around 600 wavelength both colour features overlap. Similarly, towards the final features similar observations can be made. The graph insights suggest that overlapping features are not great for determining the class because a feature value can be shared by both colours therefore reducing possibility of determining the right colour. However, a single feature that is around 520 or 640 should be good enough to determine the class. From fig.1 it is clear that those wavelength values are distinct to each colour.

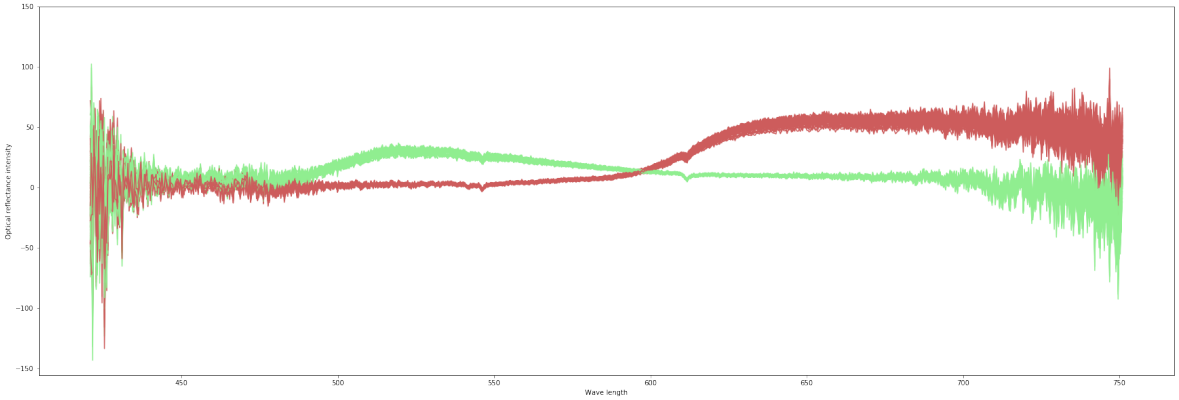

Figure 1: Input feature visualization for binary classification task. Red and green predicting features are indicated by colour.

3.2 Multi-Class Classification

Similarly, fig.2 shows different colour reflectance intensities for five different colours. Similar observation can be seen here too. However, due to larger number of colour classes some seem to overlap slightly.

From fig.**??** it can be seen that red and pink colour intensities over different wavelengths are very similar. Green and blue are also similar. Yellow on the other hand is very distinct. Just from looking at the graph it can be seen that there are not many features that

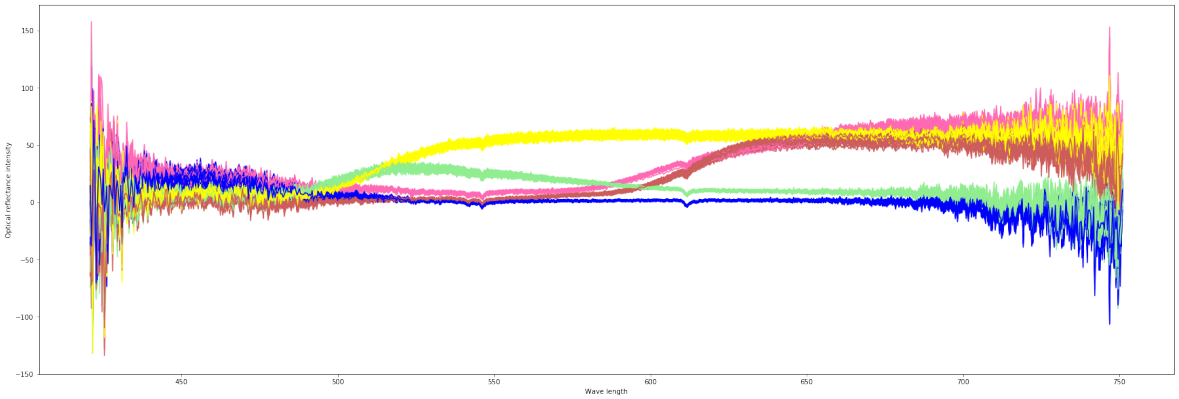

Figure 2: Input feature visualization for multi-class task. Five colour predicting features are indicated by corresponding colour.

4 Preparing Inputs and Choosing Features

5 Selecting and Training Classification Models

6 Evaluating and Comparing Model Performance

7 Result Discussion

8 Conclusion