Research Log

Rick Venema

2018-09-12

General introduction

In this research log, research is described that was carried out at the Hanze University of Applied sciences. This research is an attempt to create a classifier based on a forest data set. This dataset was used by Johnson et al. (2012) [1], their goal was to map different types of forests by using spectral data. The data set is available via the link https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping [2].

Dataset

Dataset description

This description is the description included with the site at which the data can be found, it gives a clear description on what kind of data we are dealing with.

```
This data set contains training and testing data from a remote sensing study which mapped
different forest types based on their spectral characteristics at visible-to-near infrared
wavelengths, using ASTER satellite imagery. The output (forest type map) can be used to
identify and/or quantify the ecosystem services (e.g. carbon storage, erosion protection)
provided by the forest.
```

Layout of the data

```
rawDataTesting <- read.csv("ForestTypes/testing.csv")
rawDataTraining <- read.csv("ForestTypes/training.csv")
```

The dimensions of rawDataTesting is equal to *325* and *28*. The first value represents the amount of rows and the second value represents the number of columns. The rows represent the instances, while the columns represent the attributes. The dimensions of rawDataTraining is equal to *198* and *28*.

The class attribute

The data has a class attribute, this class attribute can have different values which are presented in table 1

Table 1: Class Attributes

Abbrevation	Full name
s	'Sugi' forest
h	'Hinoki' forest
d	'Mixed deciduous' forest
o	'Other' non-forest land

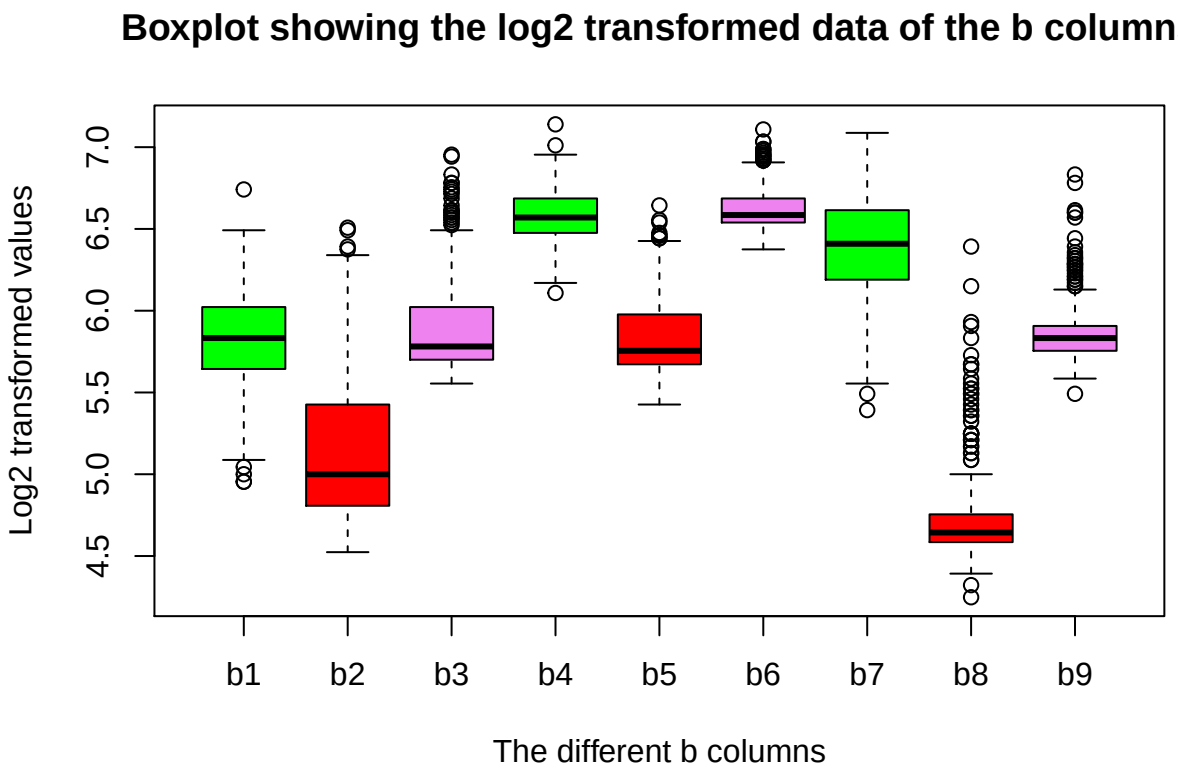

Figure 1: The log2 transformed values of the b columns.

The b columns

In figure 1 the log2 transformed data of the b columns. The boxplot give a difference in the different years, this is because the different columns can be divided into 3 different years. The first 3 boxplots represent the data from 26 September 2010, the next three boxplots represent the data from 19 March 2011, and the last three boxplots represent the data from 8 May 2011. The different colors of the boxplots represent the different spectral wavelengths. The green color represent their corresponding wavelength green(0.52-0.60 µm), red (0.63-0.69 µm) and near-infrared (NIR) (0.76-0.86 µm). NIR is represented as violet in the boxplots.

The b columns orderd by class

To get a better view of the differences between dates, the data was split into the 4 different class attributes. Each different kind of forest was put into a boxplot, which can be seen in figure 2. As can be seen, the dates differ a lot. This can be due to the different seasons at which the pictures were taken. The boxplots created from the March 2011 data, is higher than the other dates.

Research question

possible to expand this to be used in different situations. Crop recognition for example, differences in one crop.

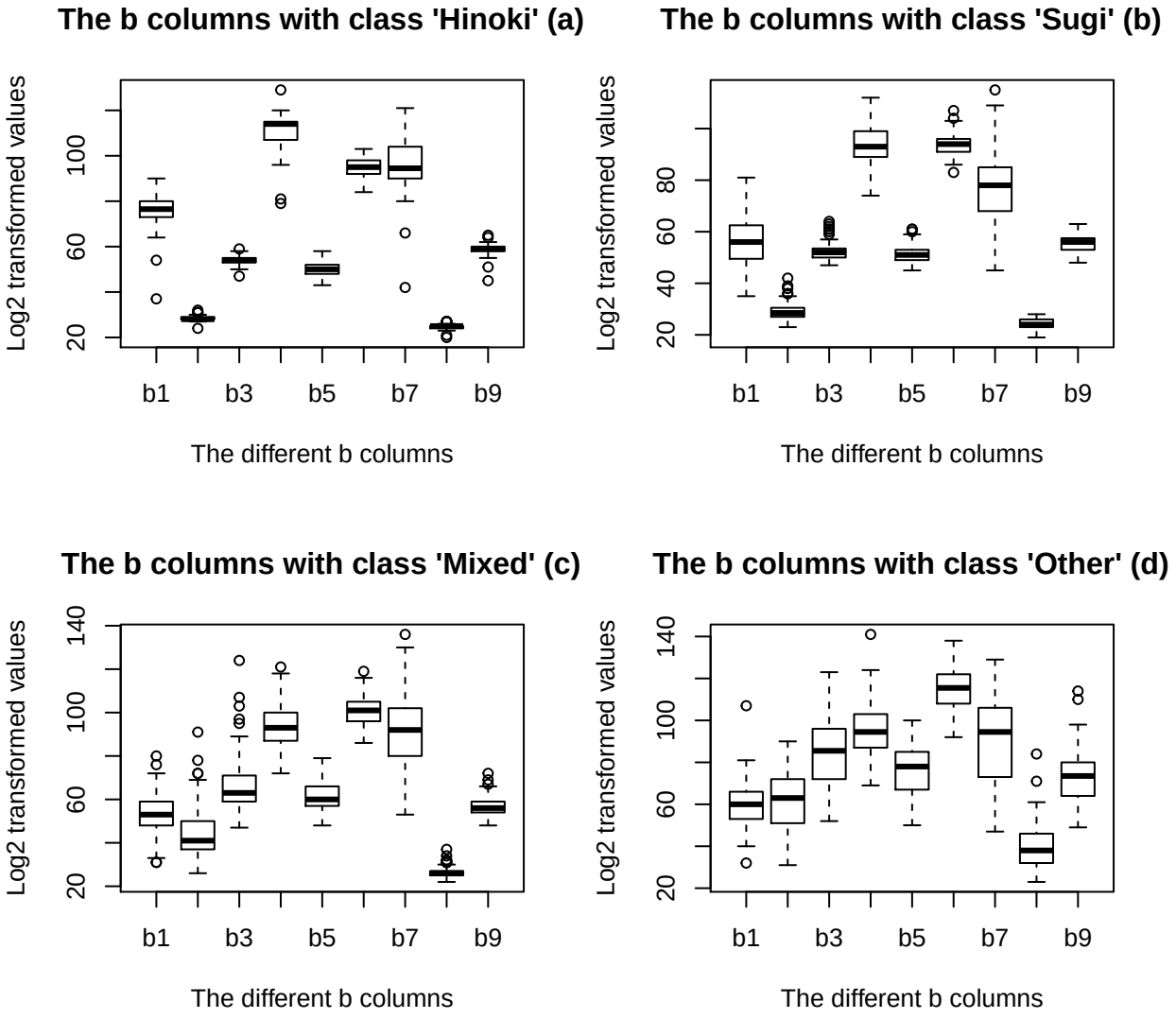

Figure 2: B columns divided by class attribute

References

[1] Johnson, B., Tateishi, R., Xie, Z., 2012. *Using geographically-weighted variables for image classification. Remote Sensing Letters*, 3 (6), 491-499.

[2] *Forest type mapping Data Set* Retrieved September 11, 2018, from https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping