Research Log

Rick Venema

2018-09-11

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

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

The next 9 columns represent ???, a summary is given

```
bColumns <- rawDataTesting[2:10]
pander(summary(bColumns))
```

Table 2: Table continues below

b1	b2	b3	b4
Min. : 31.00	Min. :23.00	Min. : 47.00	Min. : 69.00
1st Qu.: 50.00	1st Qu.:28.00	1st Qu.: 52.00	1st Qu.: 89.00
Median : 57.00	Median :32.00	Median : 55.00	Median : 95.00
Mean : 58.02	Mean :38.38	Mean : 61.47	Mean : 96.18
3rd Qu.: 65.00	3rd Qu.:43.00	3rd Qu.: 65.00	3rd Qu.:103.00
Max. :107.00	Max. :91.00	Max. :124.00	Max. :141.00

b5	b6	b7	b8	b9
Min. : 43.0	Min. : 83.0	Min. : 42.00	Min. :19.00	Min. : 45.00
1st Qu.: 51.0	1st Qu.: 93.0	1st Qu.: 73.00	1st Qu.:24.00	1st Qu.: 54.00
Median : 54.0	Median : 96.0	Median : 85.00	Median :25.00	Median : 57.00
Mean : 58.1	Mean : 99.2	Mean : 85.86	Mean :27.38	Mean : 58.88
3rd Qu.: 63.0	3rd Qu.:103.0	3rd Qu.: 98.00	3rd Qu.:27.00	3rd Qu.: 60.00
Max. :100.0	Max. :138.0	Max. :136.00	Max. :84.00	Max. :114.00

Research question

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

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