

Machine Learning Lab #2

According to the class, we know Decision Tree and K- nearest neighbor. This time we use different classifiers/regressors to analyze the data set and compare their performance.

Problem

- In this assignment you need to **use Decision Tree, K- nearest neighbor to analyze the data set.**
- You need to submit your code and report. The report should include results, using different performance metrics to analyze the results. Also you need to discuss your ideas and conclusions about the results. (e.g. You can say why a classifier is better or worse than another)

Data set

Split the data randomly to training data and test data (70% / 30%) then do your analysis

Use the Forest Fires Data Set

Attribute Information:

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMCI - FFMCI index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84

this output variable is very skewed towards 0.0, thus it may make sense to adjust the data using the logarithm.

<http://cwfis.cfs.nrcan.gc.ca/background/summary/fwi>

If you want to know what features 5-8 are you can read this website

Grading

- Post your code (20%)

- Report(80%)

For the data set with Decision Tree and K-nearest neighbor

Note, you will have to justify your choice of K

Also, you will have to justify your choice of decision tree parameters

Discuss the accuracy, specificity, and sensitivity at a minimum.

Document any parameter exploration that you thought would be helpful.