

CS4361/5361 Machine Learning

Fall 2019

Lab 4 - The `scikit-learn` library

Due Monday, October 21, 2019. Submit a paper copy by 5:00 p.m.

Email report to olacfuentes@gmail.com, include UTEP-ML2019 in the subject line.

Your task for this lab is to become familiar with the `scikit-learn` machine learning library.

Part 1

Try all the combinations of algorithms and datasets listed below, using the default parameters of the `scikit-learn` implementations. To obtain consistent results, set the seed of the random number generator to the last two digits of your student number.

Algorithms

1. k-nearest neighbors
2. Decision trees
3. Random forests
4. Logistic regression
5. Support vector machine

Datasets

1. MNIST
2. Gamma ray
3. Solar x-rays and particles

Part 2

For each dataset, find the combination of algorithm and parameter selection that yields the best results. Document your experiments and explain the meaning of the parameters of each algorithm (do not use the implementations as black boxes).

Write a report including (at least) the following items:

1. Problem description
2. Algorithms used, including parameters.
3. Experimental results. Show accuracies or mean squared errors and running times for each algorithm and parameter choice. You may want to use tables and or plots to illustrate this.
4. Discussion of results. What algorithm and parameter selection worked best? For your best-performing algorithm in each dataset, explain the parameter choice you made and the reason you believe it provided good results.
5. Conclusions.
6. Appendix: Source code