R Noah Padilla

11/3/2020

Random Forest Classification of MNIST dataset and Regression of Particles dataset

**Introduction**

The purpose of this exercise was to tune the parameters of both the random forest classifier for classifying examples in the MNIST dataset as well as the random forest regressor for the Particles dataset.

**Methods**

I performed the method of only modifying the n\_estimators or the number of trees in the random forest for both the classifier and the regressor.

**Results**

CLASSIFICATION DATA – only n\_estimators parameter was modified

|  |  |  |  |
| --- | --- | --- | --- |
| n\_estimators | Train Time | Test Time | Accuracy |
| 50 | 28.825665 secs | 0.179342 secs | 0.967000 |
| 100 | 56.486980 secs | 0.354030 secs | 0.969429 |
| 150 | 85.746600 secs | 0.523751 secs | 0.971286 |
| 200 | 114.060088 secs | 0.692670 secs | 0.971429 |
| 250 | 142.863594 secs | 0.876005 secs | 0.970571 |
| 300 | 182.422549 secs | 1.100109 secs | 0.970429 |

REGRESSION DATA – only n\_estimators parameter was modified

|  |  |  |  |
| --- | --- | --- | --- |
| n\_estimators | Train Time | Test Time | MSE |
| 10 |  |  |  |
| 20 |  |  |  |
| 30 |  |  |  |
| 40 |  |  |  |
| 50 |  |  |  |
| 60 |  |  |  |

**Conclusion**

In conclusion as the n\_estimators or number of trees in the forest increased, the accuracy and mse improved. Although they improved as the n\_estimators increased, there was a threshold where they peaked and then the accuracy and mse started to decay in performance. This is because the more generated independent trees, the better democracy is able to work.

**References**

[1] “3.2.4.3.1. sklearn.ensemble.RandomForestClassifier¶.” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. [Accessed: 03-Nov-2020].

[2] “3.2.4.3.2. sklearn.ensemble.RandomForestRegressor¶,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html. [Accessed: 03-Nov-2020].