R Noah Padilla

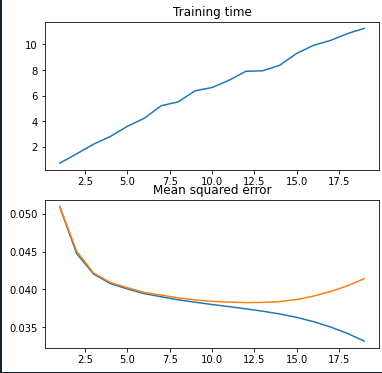
10/29/2020

Decision Tree Regressor Report

In this exercise, I reported the results of my modifications to the decision tree regressor such as the the max depth and the max leaf nodes. To find the best parameters I used the knowledge from the previous exercise to deduce which parameters were the most important – max depth and max\_leaf\_nodes. My approach to testing was to find the best depth and then to find the best mse given the best depth.

**OUTPUT WITHOUT MODIFICATIONS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Max Depth** | **Train Time** | **Time Testing Train Set** | **Train MSE** | **Time Testing Test Set** | **Test MSE** |
| 2 | 1.4613 secs | 0.0410 secs | 0.044722 | 0.0070 secs | 0.045036 |
| 4 | 2.8217 secs | 0.0610 secs | 0.040759 | 0.0100 secs | 0.040903 |
| 6 | 4.2410 secs | 0.0830 secs | 0.039447 | 0.0120 secs | 0.039603 |
| 8 | 5.5181 secs | 0.1140 secs | 0.038627 | 0.0160 secs | 0.038874 |
| 10 | 6.6390 secs | 0.1441 secs | 0.038008 | 0.0200 secs | 0.038428 |
| 12 | 7.9051 secs | 0.1790 secs | 0.037436 | 0.0240 secs | 0.038257 |
| 14 | 8.3700 secs | 0.2209 secs | 0.036764 | 0.0280 secs | 0.038393 |
| 16 | 9.9319 secs | 0.2799 secs | 0.035746 | 0.0350 secs | 0.039113 |
| 18 | 10.8428 secs | 0.3528 secs | 0.034202 | 0.0460 secs | 0.040460 |

****

Depth by Training Time Plot

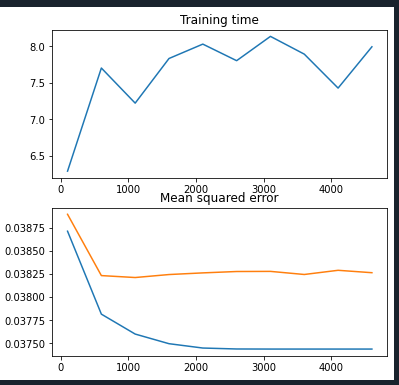
Accuracy by MSE Plot

Orange = Testing Test Set

Blue = Testing on Train Set

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Max Depth** | **Max Leaf Nodes** | **Train Time** | **Time Testing Train Set** | **Train MSE** | **Time Testing Test Set** | **Test MSE** |
| 12 | 100 | 6.2895 secs | 0.1129 secs | 0.038710 | 0.0180 secs | 0.038894 |
| 12 | 600 | 7.7017 secs | 0.1460 secs | 0.037814 | 0.0190 secs | 0.038230 |
| 12 | 1100 | 7.2210 secs | 0.1590 secs | 0.037598 | 0.0209 secs | 0.038209 |
| 12 | 1600 | 7.8321 secs | 0.1780 secs | 0.037494 | 0.0210 secs | 0.038242 |
| 12 | 2100 | 8.0291 secs | 0.1791 secs | 0.037447 | 0.0230 secs | 0.038260 |
| 12 | 2600 | 7.8023 secs | 0.1799 secs | 0.037437 | 0.0220 secs | 0.038275 |
| 12 | 3100 | 8.1350 secs | 0.1799 secs | 0.037436 | 0.0250 secs | 0.038276 |
| 12 | 3600 | 7.8920 secs | 0.1800 secs | 0.037436 | 0.0250 secs | 0.038242 |
| 12 | 4100 | 7.4270 secs | 0.1807 secs | 0.037436 | 0.0250 secs | 0.038288 |
| 12 | 4600 | 7.9923 secs | 0.1799 secs | 0.037436 | 0.0240 secs | 0.038262 |

**MODIFICATIONS TO MAX LEAF NODES**



Max Leaf Nodes by Training Time Plot

Max Leaf Nodes by MSE plot

Orange = Testing Test Set

Blue = Testing on Train Set

In conclusion, finding the optimal depth then using it with various max leaf nodes yielded a very small improvement in accuracy. To improve my findings, I don’t know what parameters would have to be modified to increase accuracy, but I do know that modifying the max\_features would increase the runtime. Although I made a slight improvement in accuracy, I lacked in the ability to find an optimal way to test parameter changes since there are many different modifications that can be made to the DecisionTreeRegressor invocation.

**References**

[1] “sklearn.tree.DecisionTreeClassifier¶.” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html. [Accessed: 27-Oct-2020].