Problem 1a

I’ve tried splits based on Entropy, Gini as well as MSE. My stopping condition is when Information gain/ Gini index/ MSE is 0, we don’t split. I’ve also varied the number of elements allowed at a Leaf (as another stopping condition).

|  |  |  |  |
| --- | --- | --- | --- |
| Len leaf node = 1 | Entropy | Gini | MSE |
| Training | 37.30998 | 35.29439 | 113.0176 |
| Test | 31.58649 | 55.36676 | 110.3071 |

|  |  |  |  |
| --- | --- | --- | --- |
| Len leaf node = 2 | Entropy | Gini | MSE |
| Training | 35.40285 | 36.12831 | 113.0176 |
| Test | 28.66182 | 50.10726 | 110.3071 |

Best Leaf node length 🡺

|  |  |  |  |
| --- | --- | --- | --- |
| Len leaf node = 5 | Entropy | Gini | MSE |
| Training | 33.69992 | 34.97282 | 115.3225 |
| Test | 23.55848 | 49.84095 | 100.4402 |

Comparison plot of predicted values on test data (leaf nodes = 5)

Problem 1b

**MSE using linear algebra exact solution for**

Training data = 24.4758827846

Test data = 24.2922381757

**MSE using Gradient descent (Stochastic) for**

Lambda used = 0.0001

I Normalized all the features X. Since for GD the value of theta (W) was not converging I used a Mean Square Error check on the training set. If the change in MSE for one iteration of GD is less than ‘Err’ (0.1, 0.01 or 0.001), then we stop applying GD.

Here’s a table of the MSE of predicted values using different values of Err.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Err = 0.1 | Err = 0.01 | Err = 0.001 |
| Training | **82.71367901** | 76.17685441 | 72.32042063 |
| Testing | **34.44922136** | 38.42811271 | 42.27111063 |

We can see that by increasing Err, we are over-fitting.