

# LEE JOO WON

## Jlee3259

### ML for Trading ML3-P1

#### REPORT

Jlee3259 | Machine Learning for Trading | 20 Feb 2017

## Methodology

After completing both RTLearner.py and BagLearner.py, I have created the test python application that runs both programs iteratively given the configuration parameters such as leaf size and number of bags. The result of program runs is displayed as tables in the reports.

For Q1, I ran the RTLearner.py iteratively to increase the number of leaf size in each loop.

Q2, I ran the BagLearner.py that uses RTLearner.py as a learner. It kept the number of bags to the fixed size but increased the size of leaf in each loop.

Q3, I ran the BagLearner.py that uses RTLearner.py as a learner also, but this time I kept the Leaf Size fixed while increasing the size of bags each loop.

## Q1: Does overfitting occur with respect to leaf size?

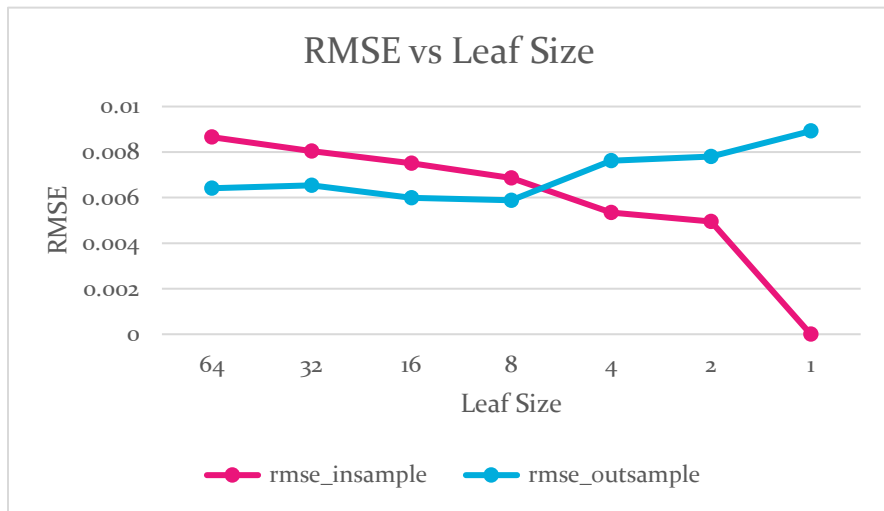
Yes, it seems pretty clear that if the size of the leaf goes below 8 the overfitting is starting to happen.

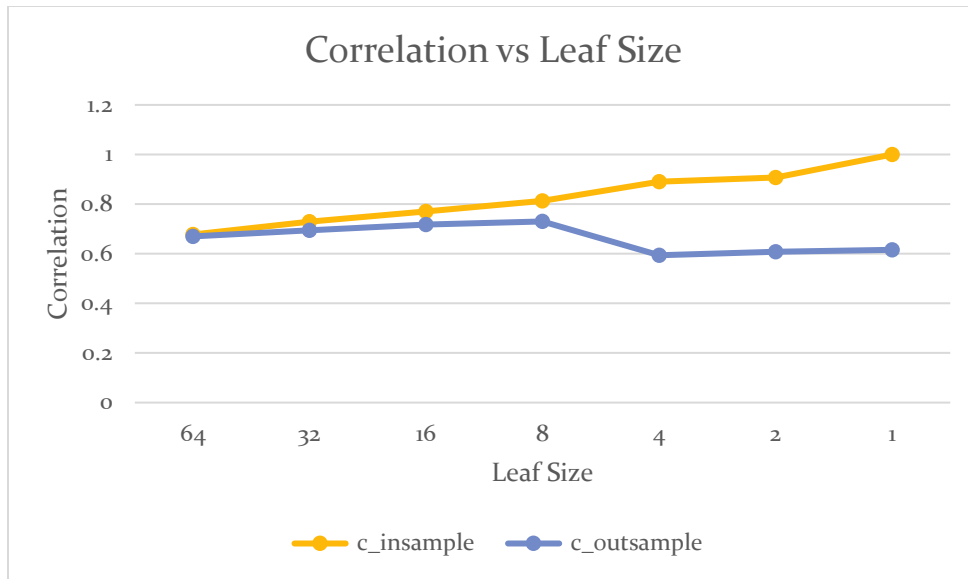
## DATA

leaf_size	rmse_insample	c_insample	rmse_outsample	c_outsample
64	0.008655422	0.67766007	0.006407304	0.66942654
32	0.008047634	0.729728643	0.006544841	0.69407147

16	0.007506427	0.770239059	0.005996676	0.717123626
8	0.006856402	0.812810393	0.005884052	0.73016937
4	0.005346926	0.890858224	0.007617843	0.593932148
2	0.004944956	0.907464154	0.00779829	0.607650701
1	0	1	0.008921616	0.615651527

## CHART





## Q2: Can bagging reduce or eliminate overfitting with respect to leaf size?

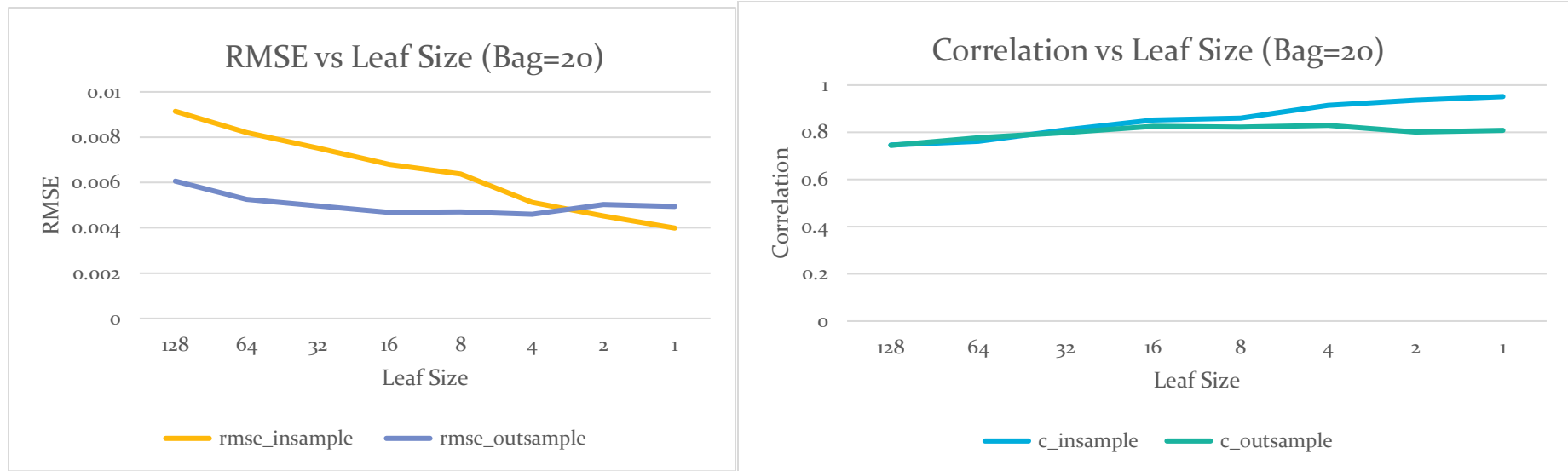
According to my data, when the number of leaf size goes lower than 4 the overfitting is starting to occur regardless of the bagging being applied.

### DATA

leaf_size	bags	rmse_insample	c_insample	rmse_outsample	c_outsample
128	20	0.009142789	0.746113008	0.006060091	0.74442285
64	20	0.008204656	0.761901281	0.005261885	0.77726387
32	20	0.007517012	0.809764718	0.004970321	0.798988522
16	20	0.00679163	0.851367827	0.004684634	0.824810553
8	20	0.00637984	0.860468928	0.004698963	0.822372563
4	20	0.005129501	0.914650025	0.004597602	0.829277527
2	20	0.004522981	0.936110623	0.005032311	0.801115193
1	20	0.003988999	0.951381738	0.004943103	0.807947957

## CHART

Chart exhibits the similar shape to the chart from the Q1 – when the experiment was executed without bagging technique.



## Q3: Does overfitting occur with respect to number of bags?

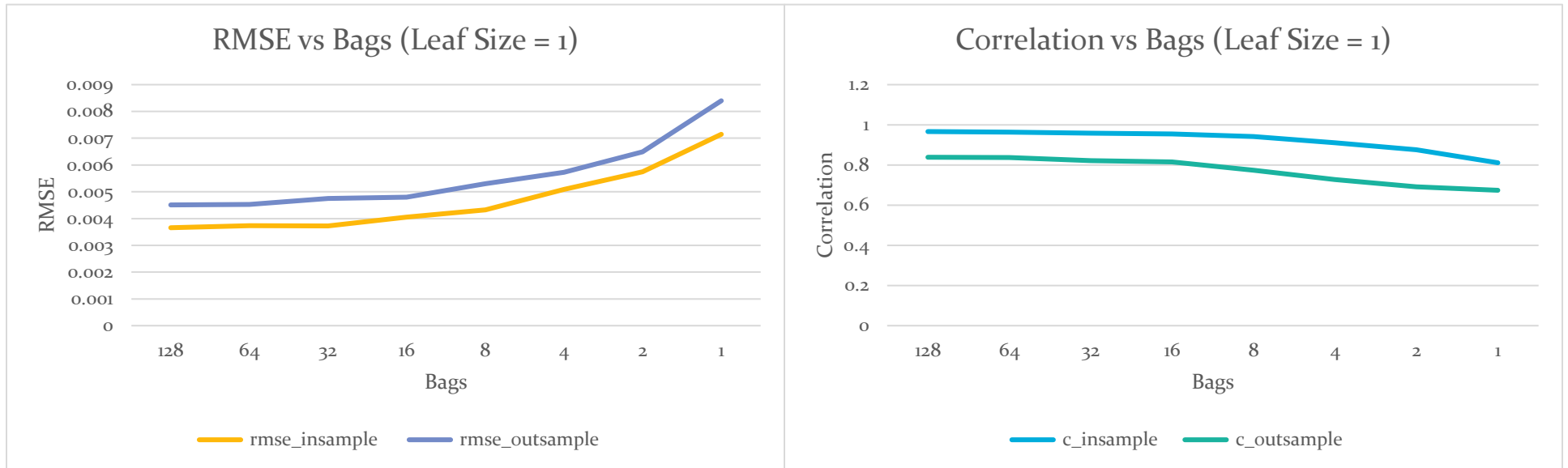
I was not able to find the symptoms of overfitting with respect to number of bags. Insample RMSE errors seems to be proportionally moving with Outsample RMSE errors.

## DATA

leaf_size	bags	rmse_insample	c_insample	rmse_outsample	c_outsample
1	128	0.003657	0.966483	0.004509	0.838527
1	64	0.003738	0.964091	0.004524	0.83702
1	32	0.003727	0.958835	0.004746	0.821018
1	16	0.00405	0.953975	0.004796	0.814799
1	8	0.004323	0.941602	0.005302	0.774279
1	4	0.005088	0.910027	0.00573	0.727144

1	2	0.005748	0.87519	0.006489	0.691531
1	1	0.007146	0.810961	0.008398	0.674062

## CHART



## DATA #2

leaf_size	bags	rmse_insample	c_insample	rmse_outsample	c_outsample
20	256	0.006929092	0.83920762	0.00476409	0.818120658
20	128	0.00689521	0.843037371	0.004637011	0.828998532
20	64	0.006973569	0.830312762	0.004911316	0.802766613
20	32	0.006876723	0.840744552	0.004770711	0.819288018
20	16	0.007016364	0.830155537	0.004909266	0.803696483
20	8	0.007625835	0.781380566	0.005175214	0.777459671
20	4	0.007694073	0.786906312	0.005394416	0.753592527
20	2	0.007592845	0.769272489	0.005848236	0.705983062
20	1	0.008295578	0.711728575	0.006457727	0.667398385

CHART #2

