Qingyang Li

Report: Project 03 – Assess learners

CS7646: ML4T - Spring 2019

Feb 10, 2019

* Does overfitting occur with respect to leaf\_size? Use the dataset istanbul.csv with DTLearner. For which values of leaf\_size does overfitting occur? Use RMSE as your metric for assessing overfitting. Support your assertion with graphs/charts. (Don't use bagging).  
  Answer:
* Can bagging reduce or eliminate overfitting with respect to leaf\_size? Again use the dataset istanbul.csv with DTLearner. To investigate this choose a fixed number of bags to use and vary leaf\_size to evaluate. Provide charts to validate your conclusions. Use RMSE as your metric.
* Quantitatively compare "classic" decision trees (DTLearner) versus random trees (RTLearner). In which ways is one method better than the other? Provide at least two quantitative measures. Important, using two similar measures that illustrate the same broader metric does not count as two. (For example, do not use two measures for accuracy.) Note for this part of the report you must conduct new experiments, don't use the results of the experiments above for this.