Weekly Report
03/01/2015

Jingyu Deng

This Week

- Rewrite the interface
    - Each tree has a global insertion order index
        - More general functions
- Modified the previous high level code
    - To maintain the last data chunk
    - Adapted to new interface

New Interface

- predict_results(self, x)
    - return a dictionary whose key-value pair is (idx, single result).
- predict_weighted_sum(self, x, weights)
    - accept a weights dictionary and return the sum result.
- def get_idx_list(self)
    - return the current index list of the ensemble.
- delete(self, idx_list)
    - accept a index list and delete trees with those indices.
- insert(self, estimators)
    - insert a list of estimators into the ensemble.
- insert_with_rf(self, n_estimators, x, y)
    - Based on (x, y), insert n_estimators random trees into the ensemble

Run Master Program

- src: https://github.com/elnio/LJrepo/tree/master/online_ensemble
- Parameters in master.py:
    - In main function:
        - n_trees: the number of trees in the ensemble
        - chunk_size: the data size for each training
        - run_simple(…)
            - run the simple version
        - run_complex(…)
            - run the complex version

Run Master Program

- Parameters in the simple version:
    - ss: the step size in each gradient descending.
    - n_test_data: the number of data points you want to test.
    - replace_flag: whether you want to replace trees during the test.
        - threshold: the replace trigger threshold
    - normalization_flag: whether you want to do normalization after each adjusting.

Run Master Program

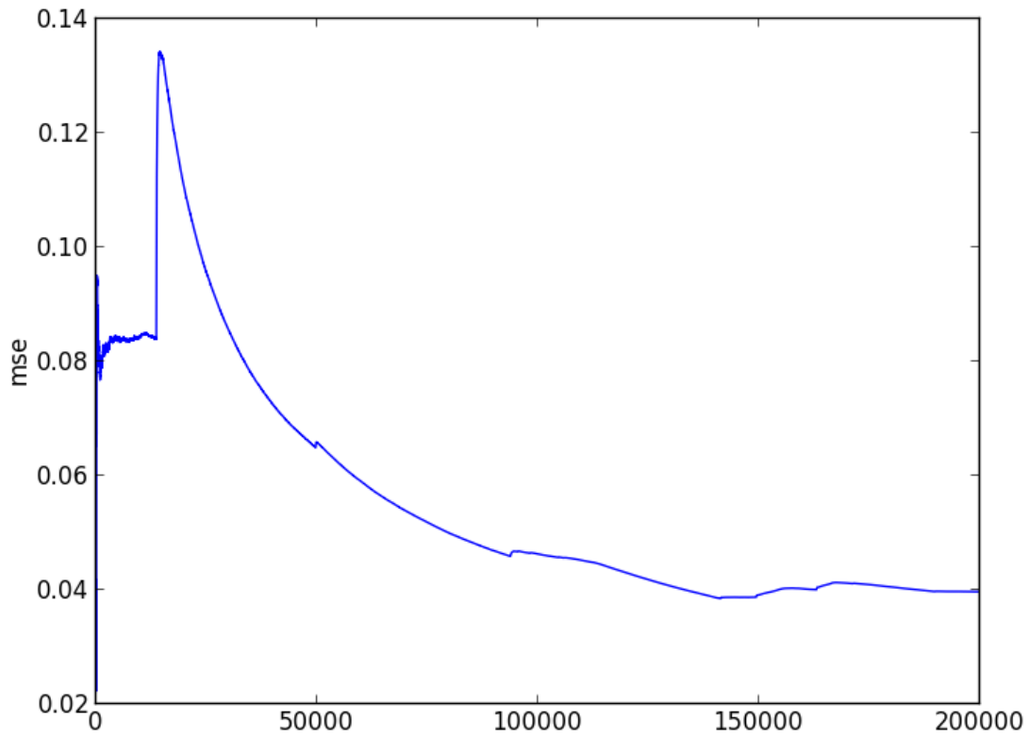
- Parameters in the complex version (besides the parameters in the simple version):
  - T: the forgetting factor

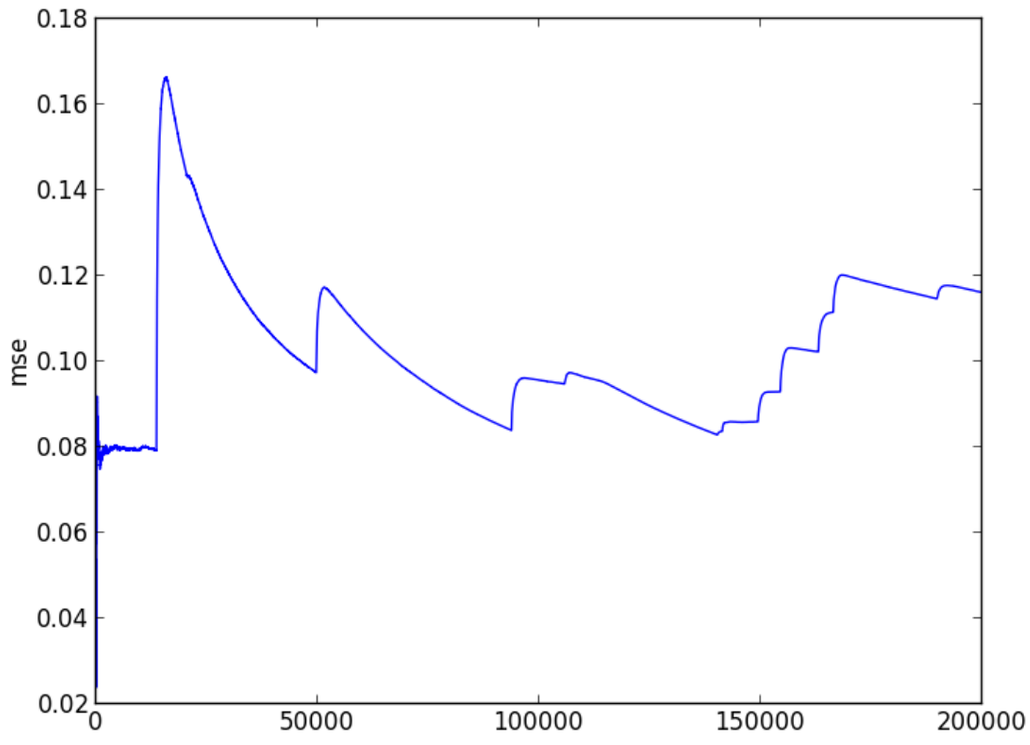
Print Messages

- After each prediction, the program will show a message such as:
  - i = 186734 predict = -0.69365 target = -0.60445 mse = 0.11568
  - index predict_value target_value current mse
- After each replace operation, the program will show a message like this:
  - replace 2 trees whose indices are [14088, 14207]

Print Message

- At the end of running, the program will draw the mse fluctuation plot such as images below:





Next Week

- Try our method on Wang's data.

- Try to reproduce Wang's result.