

1. Can H2O's GBM implementation work across multiple nodes (i.e. multiple computers, in a cluster)?

1 point

- ☒ No, it only works on a single node, and you need to use random forest for clusters.
- ☐ Yes, though very deep trees can cause communication overhead to become high.
- ☐ Yes, though it becomes unstable if you set ntrees to over 50.

2. Bagging is ... (check all that can fit)

1 point

- ☐ ...making new models that are influenced by the errors from previous models.
- ☒ ...what random forest uses.
- ☒ ...combining (e.g. averaging) the predictions of lots of models.
- ☐ ...what GBM uses.

3. Boosting is ... (check all that can fit)

1 point

- ☒ ...what GBM uses.
- ☐ ...what random forest uses.
- ☒ ...making new models that are influenced by the errors from previous models.
- ☐ ...combining (e.g. averaging) the predictions of lots of models.

4. "This change is likely to improve GBM model generalization, though you may need to increase the ntrees, to compensate."

1 point

Which of the following best fits that description?

(Hint: looking up parameters in the Appendix A of the H2O manual (

<http://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html>)

, or reading the GBM booklet (

<http://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/GBMBooklet.pdf> [PDF])

, can help you answer this question.)

- ☐ Decreasing min\_rows from 10 to 5.
  - ☐ Experimenting with nbins\_cats.
  - ☒ Decreasing learn\_rate from 0.1 to 0.02.
  - ☐ Decreasing max\_depth from 5 to 3.
  - ☐ Increasing max\_depth from 5 to 10.
-