4.1.5 Black-Box Methods

# Original

The aim is to have the proposed method be invariant to the black-box model used. For this reason, we chose to use three of the most common and high performing black-box models, which are random forests, gradient boosting, and a deep neural network (with 200 hidden layers with 200 neurons in each layer) all implemented in h2o [31]. 500 trees are used for random forests and gradient boosting, and the remaining hyperparameters kept as default. The aim is to see how well a single tree is able to reconstruct the predictions of 500 ensembles trees, or a deep network with 200 hidden layers. It is important to note these methods were not finely tuned and the relative performance of each is not of importance in this paper, as we are not trying to compare black-box methods (rather the ability to reconstruct their predictions).

# Condensed

We chose three of the most common and high performing black-box models. All are implemented in h2o [31].

* Random Forests (with 500 trees)
* Gradient Boosting (with 500 trees)
* A deep neural network (200 hidden layers with 200 neurons each)

Note: These methods were not finely tuned as their relative performance is not of importance in this paper. We are not trying to compare black-box methods we are just trying to reconstruct their predications.