

Appendix A Default hyperparameters and Descriptions

Table A1: Learning algorithms, parameters, and packages used.

| Algorithms | Parameters | Packages |
|------------|--|---|
| BG | min_samples_split = 10 max_samples = 1.0 | sklearn.ensemble.BaggingRegressor ¹ |
| DT | min_samples_split = 2 | sklearn.tree.DecisionTreeRegressor ¹ |
| MLP | hidden_layer_sizes = (100,) activation = 'relu' learning_rate_init = 0.001 momentum = 0.9 tol = 1-e4 max_iter = 200 | sklearn.neuralnetwork.MLPRegressor ¹ |
| RF | n_estimators = 100 max_features = 1.0 | sklearn.ensemble.RandomForestRegressor ¹ |
| SVM | kernel = 'rbf' gamma = scale C = 1.0 | sklearn.svm.SVR ¹ |
| XG | eta = 0.3 max_depth = 6 colsample_bytree = 1 | XGBoost ² |

¹ <https://scikit-learn.org/> - Version 1.0.2

² <https://xgboost.readthedocs.io/> - Version 0.90

Hyperparameters description:

- **BaggingRegressor:**
 - **min_samples_split:** The minimum number of samples required to split an internal node.
 - **max_samples:** The number of samples to draw from the training dataset to train each base estimator.
- **DecisionTreeRegressor:**
 - **min_samples_split:** The minimum number of samples required to split an internal node in the decision tree.
- **MLPRegressor:**
 - **hidden_layer_sizes:** The number of neurons in the hidden layers.
 - **activation:** The activation function for the hidden layers.
 - **learning_rate_init:** The initial learning rate used.
 - **momentum:** Momentum for gradient descent.
 - **tol:** Tolerance for the optimization.
 - **max_iter:** Maximum number of iterations.
- **RandomForestRegressor:**

- **n_estimators:** The number of trees in the forest.
- **max_features:** The number of features to consider when looking for the best split.
- **SVR:**
 - **kernel:** Specifies the kernel type to be used in the algorithm.
 - **gamma:** Kernel coefficient for 'rbf', 'poly', and 'sigmoid'.
 - **C:** Regularization parameter.
- **XGBoost:**
 - **eta:** Learning rate.
 - **max_depth:** Maximum depth of a tree.
 - **colsample_bytree:** The fraction of features to be randomly sampled for each tree.