
H2O.ai

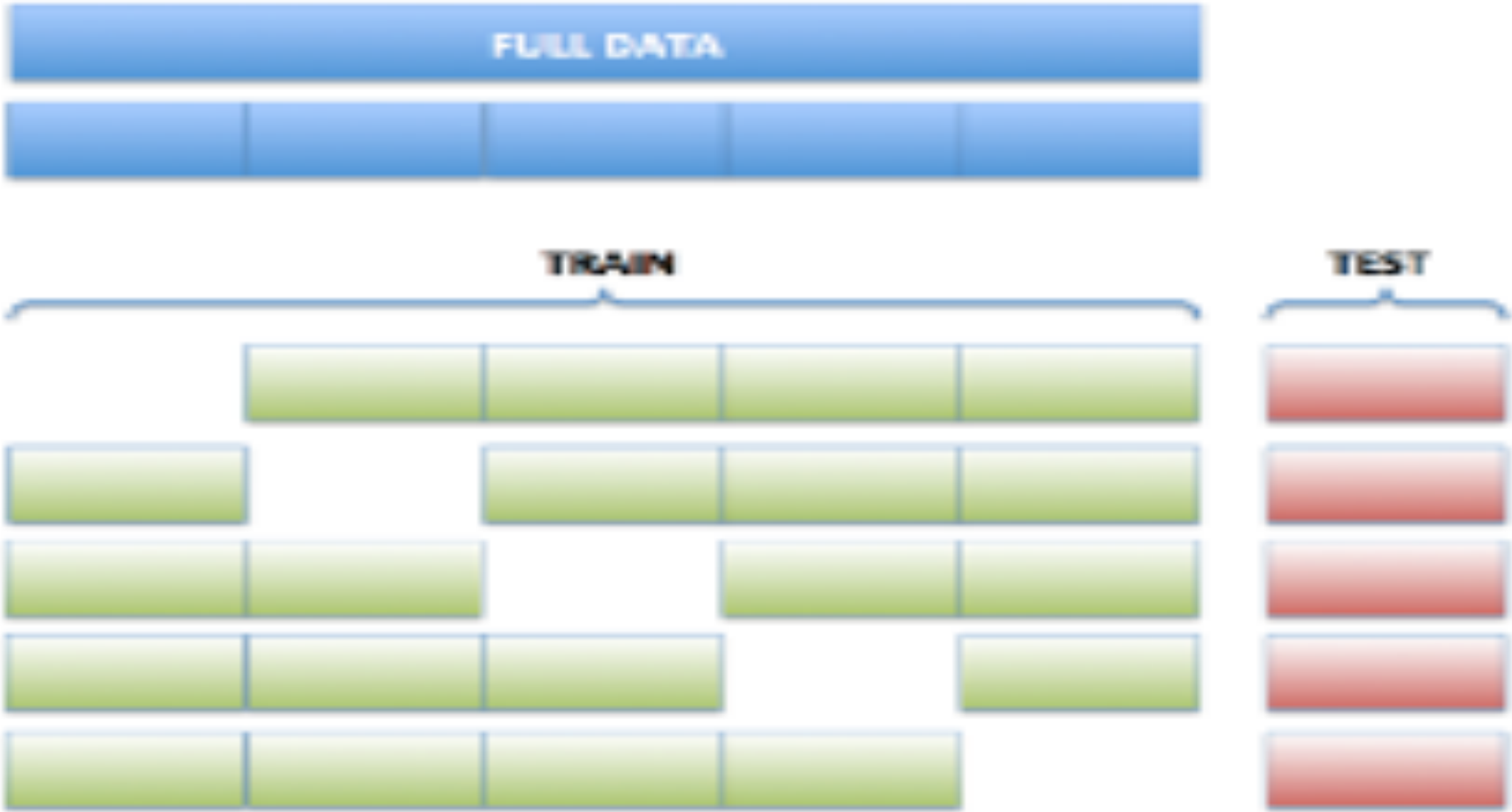
Option 2: Cross-Validation

- Partition the original data (randomly) into a training set and a test set. (e.g. 70/30)
- Train a model using the training set and evaluate performance (a single time) on the test set.

Training Set vs. Test Set

K-fold Cross-validation

- Train and test K models using separate folds.
- Average the model performance over the K test sets.
- Report cross-validated metrics.


FULL DATA
TRAIN
TEST

Validation & Cross-Validation H2O Parameters

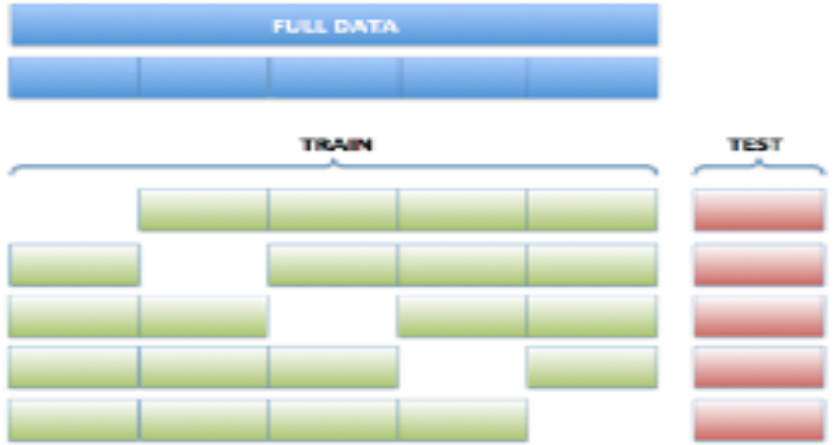
Choose one of two approaches:

1. Validation
   - **`validation_frame`**: id of the validation data frame.
2. Cross-Validation
   - **`nfolds`**: number of folds for k-fold cross-validation
   - **`fold_column`**: column with cross-validation fold index assignment per observation.
   - **`fold_assignment`**: cross-validation fold assignment scheme, if fold column is not specified.
   - **`keep_cross_validation_fold_assignment`**: keep the cross-validation fold assignment.
   - **`keep_cross_validation_predictions`**: keep the predictions of the cross-validation models.

Option 2: Cross-Validation

Training Set vs. Test Set

- Partition the original data (randomly) into a training set and a test set. (e.g. 70/30)
- Train a model using the training set and evaluate performance (a single time) on the test set.



K-fold Cross-validation

- Train and test K models using separate folds.
- Average the model performance over the K test sets.
- Report cross-validated metrics.