

## **Cross-Validation**

Use cross-validation to evaluate parameters for XGBoost.

## **Chapter Goals:**

• Learn how to cross-validate parameters in XGBoost

## A. Choosing parameters

Since there are many parameters in XGBoost and several possible values for each parameter, it is usually necessary to *tune* the parameters. In other words, we want to try out different parameter settings and see which one gives us the best results.

We can tune the parameters using cross-validation (https://en.wikipedia.org/wiki/Cross-validation\_(statistics)) (for a detailed explanation of cross-validation, see the **Data Modeling** section). In XGBoost, the cv

(https://xgboost.readthedocs.io/en/latest/python/python\_api.html#xgboost.cv) function performs cross-validation for a set of parameters on a given training dataset.

The code below demonstrates cross-validation in XGBoost.

```
1  # predefined data and labels
2  dtrain = xgb.DMatrix(data, label=labels)
3  params = {
4    'max_depth': 2,
5    'lambda': 1.5,
6    'objective':'binary:logistic'
7  }
8  cv_results = xgb.cv(params, dtrain)
9  print('CV Results:\n{}'.format(cv_results))
```

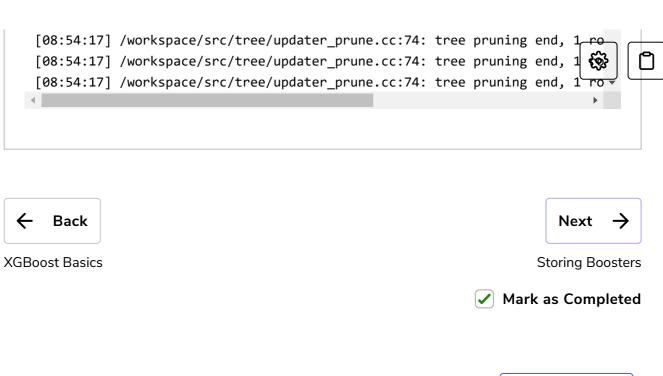
```
Output

[08:54:16] /workspace/src/tree/updater_prune.cc:74: tree pruning end, 1 ro
```

The output of cv is a pandas DataFrame (see the **Data Processing** section for details). It contains the training and testing results (mean and standard deviation) of a K-fold cross-validation applied for a given number of boosting iterations. The value of K for the K-fold cross-validation is set with the n-fold keyword argument (default is 3).

The keyword argument <code>num\_boost\_round</code> specifies the number of boosting iterations. Each boosting iteration will try to improve the model through gradient boosting. The default number of iterations is 10.

```
1 # predefined data and labels
 2 dtrain = xgb.DMatrix(data, label=labels)
 3 params = {
      'max_depth': 2,
      'lambda': 1.5,
      'objective': 'binary:logistic'
 8 cv results = xgb.cv(params, dtrain, num boost round=5)
    print('CV Results:\n{}'.format(cv_results))
 X
                                                                      2.375s
Output
 [08:54:17] /workspace/src/tree/updater_prune.cc:74: tree pruning end, 1 ro
 [08:54:17] /workspace/src/tree/updater_prune.cc:74: tree pruning end, 1 ro
```



68% completed, meet the criteria and claim your course certificate!

**Buy Certificate** 

! Report an Issue

? Ask a Question

 $(https://discuss.educative.io/tag/cross-validation\_gradient-boosting-with-xgboost\_machine-learning-for-software-engineers)\\$