Model Tuning

Each machine learning model has a set of options or configurations that can be tuned to optimize the model when fitting to data. These configurations are called **hyper-parameters**. Hence, for each hyper-parameter, there exist a range of values that can be chosen. Taking into consideration the number of hyper-parameters that an algorithm has, the entire space can become exponentially large and infeasible to explore all of them. Scikit-learn provides two convenient modules for searching through the hyper-parameter space of an algorithm to find the best values for each hyper-parameter that optimizes the model.

These modules are the

- Grid search

- Randomized search

Grid Search

Grid search comprehensively explores all the specified hyper-parameter values for an estimator. It is implemented using the **GridSearchCV** module. Let's see an example using the Random forest for regression. The hyper-parameters we'll search over are

- The number of trees in the forest, **n_estimators**

- The maximum depth of the tree, **max_depth**

- The minimum number of samples required to split an internal node, **min_samples_leaf**

```
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn import datasets

# load dataset
data = datasets.load_boston()

# separate features and target
X = data.data
y = data.target
```