**Section 6.7: Tree Based Methods in Python**

**Duration:** 2.5 hours

**Concepts:**

* Regression trees
* Classification trees
* Bagging
* Random forests
* Boosting
* Bayesian additive regression trees

**Textbook section:** An Introduction to Statistical Learning in Python, Chapter 8

|  |  |
| --- | --- |
| **Materials and Resources** | **Learning Goals** |
| * Computers for students with Jupyter Notebook * Tree Based Methods Slides * Tree Based Methods Exercises Jupyter Notebook file | * Decision trees for regression and classification * Several methods for improving the prediction accuracy of trees |

|  |  |  |
| --- | --- | --- |
| **Duration** | **Lesson Section** | **Learning Objectives** |
| 40 mins | Go through the regression trees and classification trees sections. | * What is a tree? * Algorithm to build trees * Recursive binary splitting * Cost complexity pruning * Trees for classification * Trees for regression |
| 25 mins | Go through the “Fitting Classification Trees” and “Fitting Regression Trees” sections in the Jupyter Notebook file as a class. | * Fitting regression and classification trees with `DecisionTreeClassifier()` * Plotting trees * Pruning trees with `cost\_complexity\_pruning\_path()` * Estimating test error rate |
| 15 mins | Go through the bagging and random forests sections. | * Bagging method for regression and classification trees * Out-of-bag error estimation * Random forests method for regression and classification trees |
| 10 mins | Go through the “Bagging and Random Forests” sections in the Jupyter Notebook file as a class. | * Fit bagged regression trees and random forests with `RandomForestRegressor()` * Use `feature\_importances\_` to measure relative importance of variables |
| 15 mins | Go through the boosting section. | * The boosting method for regression trees * Important parameters |
| 15 mins | Go through the “Boosting” section in the Jupyter Notebook file as a class. | * Fit boosted regression trees with `GradientBoostingRegressor()` * Look at the relative importance of variables |
| 15 mins | Go through the Bayesian Additive Regression Trees section. | * The BART method for regression trees |
| 8 mins | Go through the “Bayesian Additive Regression Trees” section in the Jupyter Notebook file as a class. | * Fit BART using `BART()` * Compare test errors between methods |