

Regression: Tree-based Methods

1. Decision Tree (DT)

- Choose a range for the maximum tree depth.
 - Start with a low depth value (e.g., 1) and finish with `None` (deepest tree possible). Set other parameters as default.
 - Plot a chart with training and validation scores based on these different values.
 - What do you think is the best maximum depth?
- Perform the same study with the “minimum samples split”.
- Perform the same study with the “minimum samples leaf”.
- Compare performance when using different `criterion`.
- Train a DT with low depth (e.g., 3 or 4) and plot it.

2. Random Forest (RF)

- Choose a range for the number of estimators.
 - Start with a low number (e.g., 1) and finish with 2,000. Set other parameters as default.
 - Plot a chart with training and validation scores based on these different values.
 - What do you think is the best number of estimators?
- Perform the same study with the number of maximum features. Try options like "sqrt", "log2", or even a fraction of the total features.
- Experiment with using bootstrap samples (default) vs. the whole dataset for building each tree.
- Analyze the importance of each feature. Which are the most important ones? And the least important?

3. Gradient Boosted Decision Trees (GBDT)

- Choose a range for the learning rate.
 - Start with a low value (e.g., 0.001) and finish with a big one (e.g., 100). Set other parameters as default.
 - Plot a chart with training and validation scores based on these different values.
 - What do you think is the best learning rate?
- Perform the same study with the tree depth.
- Experiment with early stopping.

4. Hyperparameter Tuning

- Find good hyperparameters for the tree-based methods and kNN.
- Use `RandomizedSearchCV` and `GridSearchCV` from `sklearn`.