Apr 17, 2023

Machine Learning Methods and Applications

Week 6. Decision trees - II

Remember

- Decision trees are interpretable, easy to understand, and non-linear models.
- They can handle missing values and robust to outliers.
- They may overfit easily.
- Hyperparameters can be used to control the learning rate of a decision tree.

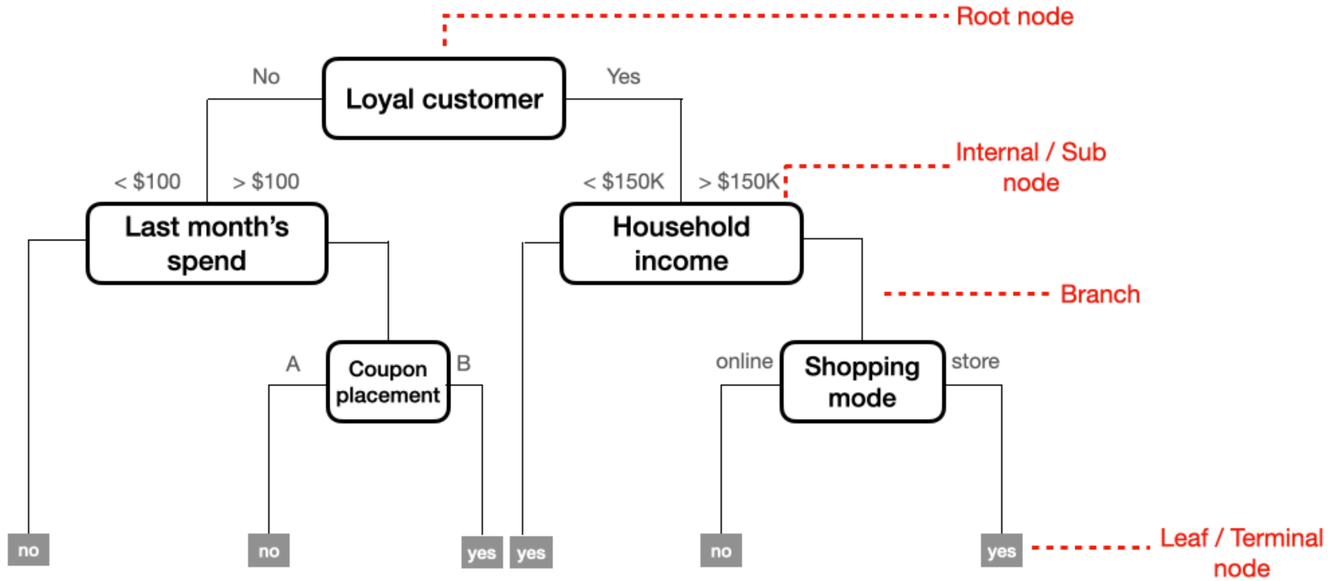
Parameters vs. hyperparameters

Model **parameters** are being fit during training; they are the result of model training. However, **hyperparameters** are being set before training.

Decision trees’ hyperparameters

Hyperparameters of a decision tree

- cp
- minsplit
- minbucket
- maxdepth
- …



The name of the hyperparameters may used differently in other packages and languages.

Cost complexity (cp)

It is used to control the size of the decision tree and to select the optimal tree size.

- It takes the any value between 0 and 1.

- Higher **cp**, less deep tree

Minimum split (minsplit)

The minimum number of observations that must exist in a node in order for a split to be attempted.

Lower **minsplit**, more deep tree

Minimum bucket (minbucket)

The minimum number of observations in any terminal node. In most of the tools, the value of **minbucket** is setted according to the value of **minsplit**.

Lower **minbucket**, more deep tree

Maximum depth (maxdepth)

The maximum number of node levels in a tree.

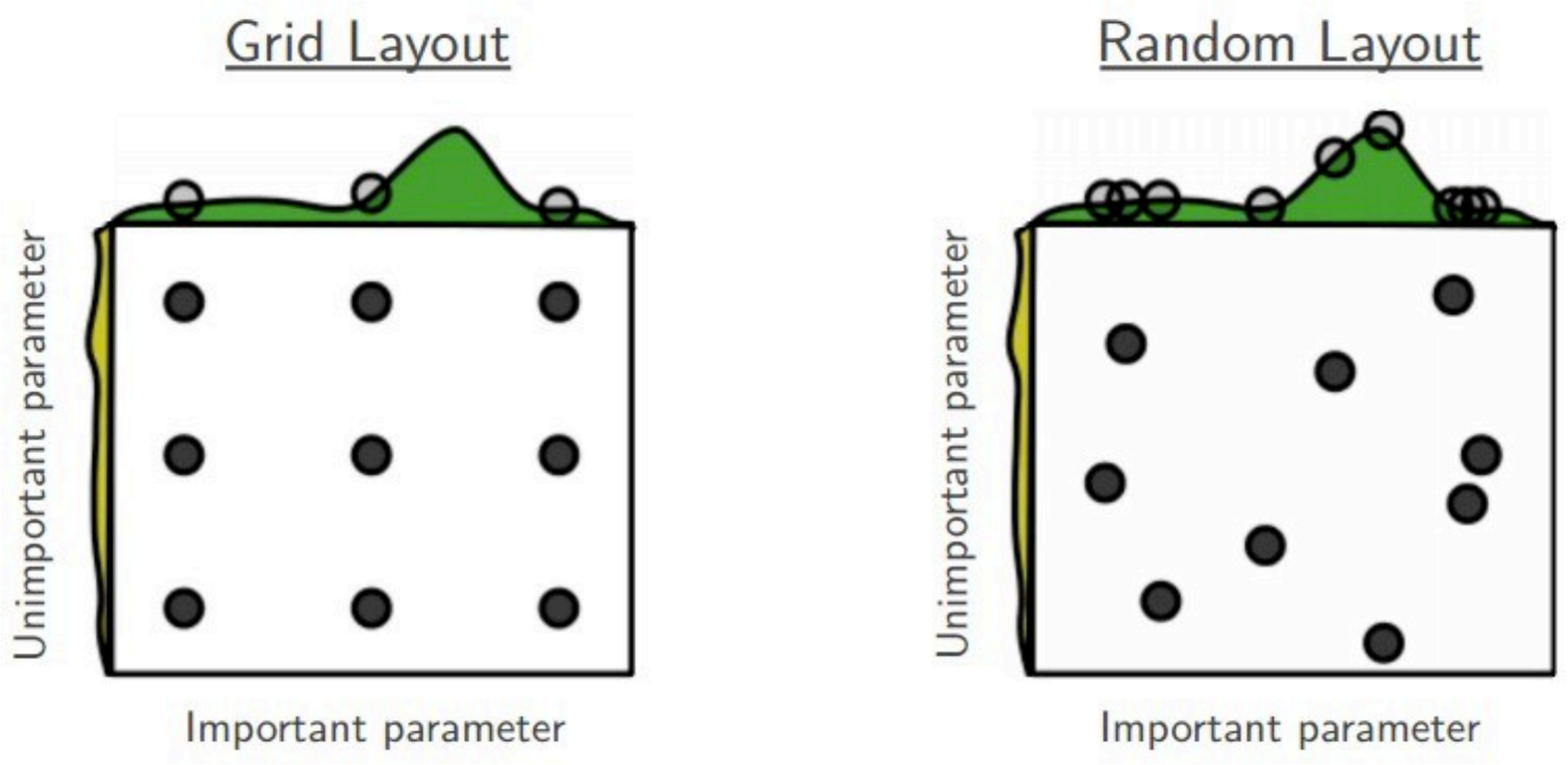
Lower **maxdepth**, less deep tree

Validation set

Validation set

Measuring model performance on the test data during hyperparameter tuning may give a biased estimation.

Hyperparameter search methods



Credit: Bergstra and Bengio (2012) JMLR 2012

Grid vs. random search

- Grid search can get slow and computationally expensive.
- Grid search computes the all hyperparameter combinations.
- Random search computes the random subsets of hyperparameter combinations.

Application

See the R codes on the course GitHub repository!

The video recording of today's lecture will be available on YouTube, and slides on **GitHub**.
Feel free to contact me via e-mail: mustafacavus@eskisehir.edu.tr