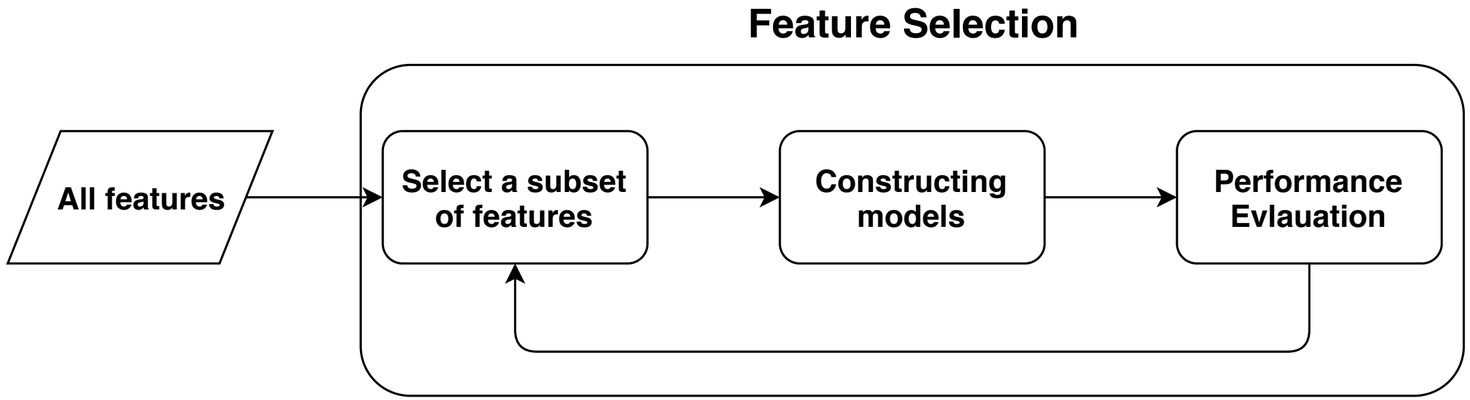
DLI Accelerated Data Science Teaching Kit

Lecture 3.5 - Feature Selection: Introduction to Model-based Methods

Model-based Methods for Feature Selection

Model-based methods learn the best features that contribute to the model performance when constructing these models.



- The most common methods are regularization methods.

- It is to penalize the model, which introduce additional constraints into the optimization of learning the models with lower complexity.

LASSO (Least Absolute Shrinkage and Selection Operator)

It is a powerful method that perform two main tasks: regularization and feature selection.

- First formulated by Robert Tibshirani in 1996.
- The LASSO method puts a constraint on the sum of the absolute values of the model parameters, the sum has to be less than a predefined value (upper bound).
- In order to do so the method apply a shrinking (regularization) process where it penalizes the coefficients of the regression variables shrinking some of them to zero.
- During features selection process the variables that still have a non-zero coefficient after the shrinking process are selected to be part of the model. The goal of this process is to minimize the prediction error.

LASSO (Least Absolute Shrinkage and Selection Operator)

There are many advantages in using LASSO method.

- It can provide a promising prediction accuracy, because shrinking and removing the coefficients can reduce variance without a substantial increase of the bias.
- This is especially useful when the data has a small number of samples and a large number of features.
- LASSO helps increase the model interpretability by eliminating irrelevant variables that are not related to the target variable, which would reduce overfitting as well.

LASSO in the Linear Model

Linear model, often called *Linear Regression Model*, describes the relationship between response *Y* and explanatory variables X.

- The case of one explanatory variable is called Simple Linear Regression while the case with two or more explanatory variables is called Multiple Linear Regression.

- An assumption is the linearity of the model, that is a linear relationship between the response variable and the explanatory variables.

$$Y = w_0 + w_1 \times x_1 + \cdots + w_n \times x_n = W \times X$$

where the parameters $w_0, w_1, \ldots, w_n$ are the weights for the model and n is the number of the explanatory variables (features). Y is the response variable (target variable).

LASSO Based Object Function for Linear Models

The objective of optimizing the model is to minimize the mean square errors.

$$minimize\left(\frac{\|Y - W\times X\|_2^2}{n}\right) \qquad subject\ to \qquad \|W\|_1 < t$$

where *t* is the upper bound for the sum of the weights, n is the number of samples.

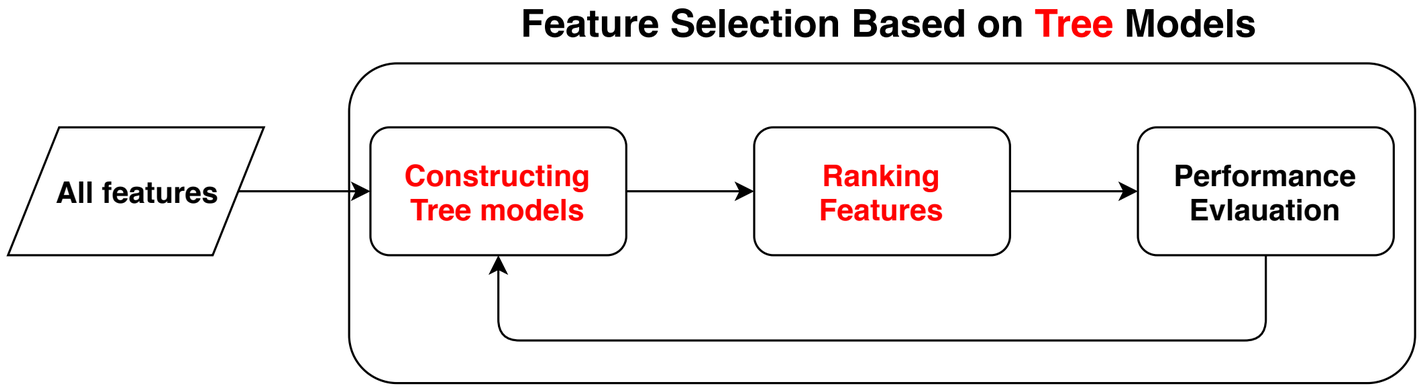
This optimization problem is equivalent to the parameter estimation that follows.

$$argmin\left(\frac{\|Y - W\times X\|_2^2}{n} + \lambda\times\|W\|_1\right)$$

where $\|W\|_1 = \sum_{i=1}^{n}|w_i|$. $\lambda \geq 0$ that controls the strength of the penalty. The larger the value of λ is, the greater the amount of shrinkage is.

Feature Selection via Tree Model

Tree-based methods will estimate the feature importance during the procedure of constructing models to select important features.



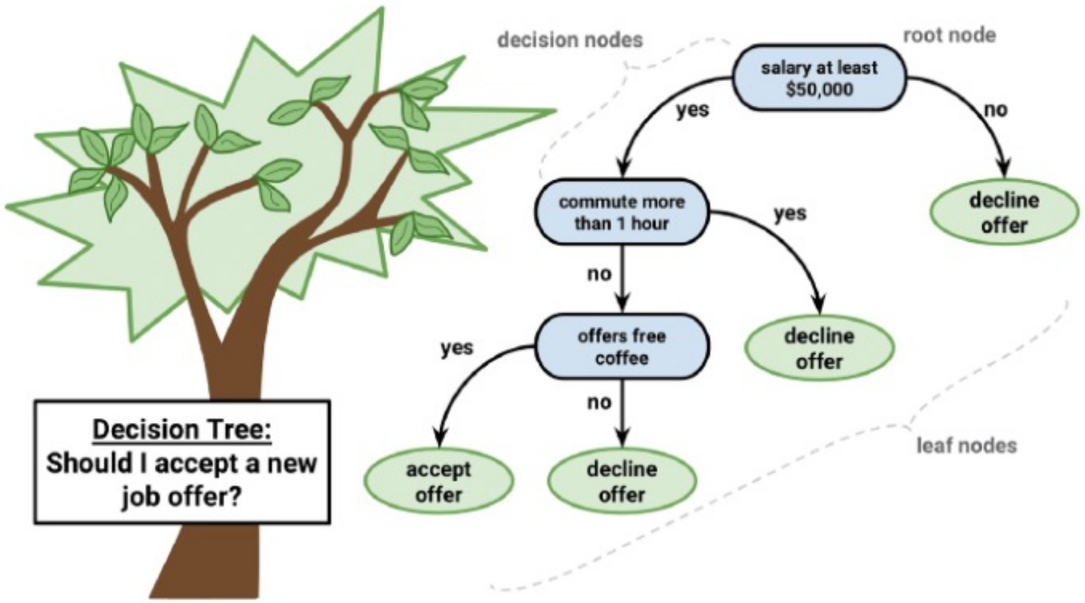
- The most common method could be decision tree and random forest.

Decision Tree

Decision tree builds classification or regression models in the form of a tree structure.

- A decision tree predicts the value of a target variable by following the decisions in the tree from the root (beginning) down to a leaf node.

- A tree consists of branching conditions where the value of a predictor is compared to a trained weight.
    - The number of branches and the values of weights are determined in the training process.



Source: https://medium.com/analytics-vidhya/machine-learning-decision-tree-regression-ff8563ffaf52

DEEP
LEARNING
INSTITUTE

PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

Thank You