## Gradient Boosting Classifier - Overview

Gradient Boosting builds an additive ensemble of shallow trees by sequentially fitting each tree to the negative gradient (residuals) of a differentiable loss. Learning rate and tree depth control model complexity and generalization.

This example uses Gradient Boosting (scikit-learn via reticulate) to classify the Iris dataset. Workflow: split train/test, train, predict, and evaluate (classification metrics).

Prerequisites - R packages: daltoolbox, daltoolboxdp - Python accessible via reticulate (scikit-learn installed)

# Gradient Boosting Classifier  
  
# Installation (if needed)  
#install.packages("daltoolboxdp")

# Loading packages  
library(daltoolbox)  
library(daltoolboxdp)

# Loading Iris dataset  
iris <- datasets::iris

# Training and evaluation with Gradient Boosting  
  
slevels <- levels(iris$Species) # target variable levels  
  
set.seed(1)  
sr <- sample\_random() # stratified random sampling  
sr <- train\_test(sr, iris) # split data  
iris\_train <- sr$train  
iris\_test <- sr$test  
  
# Numeric encoding of the target for scikit-learn (keeping Species as original target)  
iris\_train$species\_encoded <- as.integer(as.factor(iris\_train$Species))  
iris\_train\_label <- iris\_train[, !names(iris\_train) %in% "Species"]  
  
# 1) Train  
model <- skcla\_gb("species\_encoded", slevels)  
model <- fit(model, iris\_train\_label)  
train\_prediction <- predict(model, iris\_train\_label)  
  
# 2) Evaluate on train  
iris\_train\_predictand <- adjust\_class\_label(iris\_train[, "Species"]) # original labels  
train\_eval <- evaluate(model, iris\_train\_predictand, train\_prediction)  
print(train\_eval$metrics)

## accuracy TP TN FP FN precision recall sensitivity specificity f1  
## 1 1 39 81 0 0 1 1 1 1 1

# 3) Evaluate on test  
iris\_test$species\_encoded <- as.integer(as.factor(iris\_test$Species))  
iris\_test\_label <- iris\_test[, !names(iris\_test) %in% "Species"]  
test\_prediction <- predict(model, iris\_test\_label)  
  
iris\_test\_predictand <- adjust\_class\_label(iris\_test[, "Species"])  
test\_eval <- evaluate(model, iris\_test\_predictand, test\_prediction)  
print(test\_eval$metrics)

## accuracy TP TN FP FN precision recall sensitivity specificity f1  
## 1 0.9333333 11 19 0 0 1 1 1 1 1

References - Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.