About the method - cla\_knn: k-Nearest Neighbors classifier. Classifies by majority vote among the k nearest neighbors in feature space. - Main parameter: k (number of neighbors). Small k may overfit noise; larger k smooths the decision boundary.

Environment setup: install and load the package.

# Classification using KNN  
  
# installation   
#install.packages("daltoolbox")  
  
# loading DAL  
library(daltoolbox)

Load sample data (iris) and initial inspection.

iris <- datasets::iris  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

Identify target Species levels to configure the classifier.

# extracting the levels for the dataset  
slevels <- levels(iris$Species)  
slevels

## [1] "setosa" "versicolor" "virginica"

Building train and test samples via random sampling Train/test split with reproducible random sampling.

# Building train and test samples via random sampling  
set.seed(1)  
sr <- sample\_random()  
sr <- train\_test(sr, iris)  
iris\_train <- sr$train  
iris\_test <- sr$test

Check class distribution for full dataset, training, and test.

tbl <- rbind(table(iris[,"Species"]),   
 table(iris\_train[,"Species"]),   
 table(iris\_test[,"Species"]))  
rownames(tbl) <- c("dataset", "training", "test")  
head(tbl)

## setosa versicolor virginica  
## dataset 50 50 50  
## training 39 38 43  
## test 11 12 7

Model training Train the kNN model: set target, levels, and k.

# Model training  
model <- cla\_knn("Species", slevels, k=1) # k=1 for nearest neighbor  
model <- fit(model, iris\_train)

Training evaluation Predict on training set, adjust labels, and compute metrics.

# Checking fit on training data  
train\_prediction <- predict(model, iris\_train)  
  
# Model evaluation (training)  
iris\_train\_predictand <- adjust\_class\_label(iris\_train[,"Species"])  
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

Test evaluation Predict on test set and compute metrics.

# Model test  
test\_prediction <- predict(model, iris\_test)  
  
iris\_test\_predictand <- adjust\_class\_label(iris\_test[,"Species"])  
  
# Test evaluation  
 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 - Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory.