About the utility - clu\_tune: selects hyperparameters for a clustering method. In this example, it chooses k for cluster\_kmeans over a range.

# Clustering - Tune Kmeans  
  
# installation   
#install.packages("daltoolbox")  
  
# loading DAL  
library(daltoolbox)

Load data (iris).

data(iris)

Fit the model with a search over k=1..10 and extract the best k.

# model training with hyperparameter search  
model <- clu\_tune(cluster\_kmeans(k = 0), ranges = list(k = 1:10))  
model <- fit(model, iris[,1:4])  
model$k

## [1] 9

Generate cluster labels with the best k.

# run with best parameter  
clu <- cluster(model, iris[,1:4])  
table(clu)

## clu  
## 1 2 3 4 5 6 7 8 9   
## 4 11 50 11 11 12 11 18 22

External evaluation with Species.

# external evaluation using ground truth labels  
eval <- evaluate(model, clu, iris$Species)  
eval

## $clusters\_entropy  
## # A tibble: 9 × 4  
## x ce qtd ceg  
## <fct> <dbl> <int> <dbl>  
## 1 1 0 4 0   
## 2 2 0 11 0   
## 3 3 0 50 0   
## 4 4 0 11 0   
## 5 5 0 11 0   
## 6 6 0 12 0   
## 7 7 0.439 11 0.0322  
## 8 8 0.650 18 0.0780  
## 9 9 0 22 0   
##   
## $clustering\_entropy  
## [1] 0.1102325  
##   
## $data\_entropy  
## [1] 1.584963

References - Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior.