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

About the technique - smoothing\_inter: discretization/smoothing by regular intervals (equal widths). Useful to summarize continuous variables into ranges.

Sample data and general idea of discretization/smoothing.

# Discretization and smoothing  
# Discretization: transform continuous functions, models, variables, and equations into discrete versions.   
  
# Smoothing: create an approximating function to capture important patterns, reducing noise and high-frequency variation.  
  
# Defining bin intervals is essential to enable the approximation/discretization.  
  
# General function to evaluate different smoothing techniques  
  
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

Apply interval-based discretization and inspect bins.

# smoothing using regular interval  
obj <- smoothing\_inter(n = 2)   
obj <- fit(obj, iris$Sepal.Length)  
sl.bi <- transform(obj, iris$Sepal.Length)  
print(table(sl.bi))

## sl.bi  
## 5.32842105263158 6.73272727272727   
## 95 55

obj$interval

## [1] 4.3 6.1 7.9

Evaluate conditional entropy between bins and species.

entro <- evaluate(obj, as.factor(names(sl.bi)), iris$Species)  
print(entro$entropy)

## [1] 1.191734

Optimize the number of bins (search 1:20) and apply again.

# Optimizing the number of binnings  
  
opt\_obj <- smoothing\_inter(n=1:20)  
obj <- fit(opt\_obj, iris$Sepal.Length)  
obj$n

## [1] 8

obj <- fit(obj, iris$Sepal.Length)  
sl.bi <- transform(obj, iris$Sepal.Length)  
print(table(sl.bi))

## sl.bi  
## 4.52727272727273 5.00294117647059 5.49 5.88333333333333   
## 11 34 20 30   
## 6.352 6.76666666666667 7.23333333333333 7.71666666666667   
## 25 18 6 6