2104529 Computational Methods in Industrial Engineering

Workshop 5: Advance Data Exploration and Classification

Question 1

Consider `iris` dataset, develop non-tree based methods (a.k.a non-rule based) to classify data and discuss its limitation.

(a) stratify 10 data points of each Species using `set.seed(17)` .

(b) develop knn model and show that it is related to Voronoi diagram

(c) develop ann model and show that it is related to linear regression

(d) develop svm model and explain tuning parameters

Question 2

Consider `iris` dataset (again), develop tree based methods to classify data and discuss its limitation.

(a) develop decision tree model with `rpart::rpart()` and explain Complexity Parameter (cp)

(b) develop decision tree model with `party::ctree()` and explain the difference from the previous model

(c) develop random forest model with `randomForest::randomForest()` and explain tuning parameters

Question 3

Consider dataset '`MASS::Cars93`' package and answer the following questions

(a) Find and remove data that are *not avaiable* ('`NA`')

(b) Does '`non-USA`' cars and '`USA`' cars of the same '`Type`' difference from each other, in terms of MPG, Price, Engine Size, Dimension ? If so, how?

(c) Does '`Origin`' and '`Type`' of vehicle affect MPG in the city and in the highway?

(d) Is these any variation between '`Origin`' and '`Type of vehicle`'?

(e)

Question 4

Consider an example of tax cheating profiling (similar to lecture), illustrate how decision tree and information gain works.

(a) develop decision tree model with `party::ctree()`

(b) use concept of information gain with `funModel::information_gain()`

Question 5

A distribute center has stored a fast-moving consumer good (FMCG) and operated by its logistics service provider. It recently found that the recommendation of TiHi (Tire-Height) of one product group is incorrect as the DC allows a slightly more overhang on the pallet. As a result, its experimented with the number of cartons stacking on a pallet. Based on the sales data, current TiHi of item, and the experimental data, explore and classify numbers of layer per pallet and number of carton per layer.

last updated: October 25, 2022