

R notes

Notes, tips and tricks that I'm picking up as I do R

factors

Way of defining categorical data in R e.g “West” “East” or “Male” “Female”

Pass in a vector to the *factor* function to create a factor

Data Frame

Table in R. Essentially a 2D array where the column represents the variable and the row is a unique entry in the table

attach()

Allows the variables in a data frame (columns) to be referenced in code without the need for syntax such as `data_frame_name$column_name`

Columns can now just be referenced by name

Not advisable to use `attach` is working with multiple data frames/datasets

rpart

complexity parameter (cp): The minimum improvement required in the model for a split to take place. Based on the cost of the complexity model defined as

$$\sum_{Terminalnodes} Misclass_i + \lambda * (Splits)$$