Decision Tree

Kiya Govek

1/16/2018

All of this is probably broken from moving on to Python. Just putting it here as an archive of what we tried in R with decision trees in case we want to come back

Read in data

```
setwd("~/Documents/Carleton/Comps/R_code")
evans <- read.csv("data/Evans_Room_Temp_and_Valve.csv")
names(evans)
```

```
##   [1] "Date"          "Time"          "EV.CUH2.V"     "EV.CUH3.V"
##   [5] "EV.CUHA1.V"    "EV.CUHB1.V"    "EV.HX1.V"      "EV.RM003.V"
##   [9] "EV.RM102.V"    "EV.RM106.V"    "EV.RM107.V"    "EV.RM108.V"
##  [13] "EV.RM109.V"    "EV.RM111.V"    "EV.RM112.V"    "EV.RM114.V"
##  [17] "EV.RM116.V"    "EV.RM118.V"    "EV.RM119.V"    "EV.RM120.V"
##  [21] "EV.RM121.V"    "EV.RM122.V"    "EV.RM200.V"    "EV.RM202.V"
##  [25] "EV.RM203.V"    "EV.RM204.V"    "EV.RM205.V"    "EV.RM206.V"
##  [29] "EV.RM207.V"    "EV.RM211.V"    "EV.RM212.V"    "EV.RM213.V"
##  [33] "EV.RM214.V"    "EV.RM215.V"    "EV.RM216.V"    "EV.RM300.V"
##  [37] "EV.RM302.V"    "EV.RM303.V"    "EV.RM304.V"    "EV.RM305.V"
##  [41] "EV.RM306.V"    "EV.RM307.V"    "EV.RM311.V"    "EV.RM312.V"
##  [45] "EV.RM313.V"    "EV.RM314.V"    "EV.RM315.V"    "EV.RM316.V"
##  [49] "EV.RM400.V"    "EV.RM401.V"    "EV.RM402.V"    "EV.RM403.V"
##  [53] "EV.RM404.V"    "EV.RM405.V"    "EV.RM407.V"    "EV.RM411.V"
##  [57] "EV.RM412.V"    "EV.RM413.V"    "EV.RM414.V"    "EV.RM415.V"
##  [61] "EV.RMB4.V"     "EV.RMG04.V"    "EV.RMG05.V"    "EV.RMG06.V"
##  [65] "EV.RMG07.V"    "EV.RMG08.V"    "EV.RMG09.V"    "EV.RMG10.V"
##  [69] "EV.RMG11.V"    "EV.RMG14.V"    "EV.RMG16.V"    "EV.RMG17.V"
##  [73] "EV.RMG18.V"    "EV.RMG19.V"    "EV.RMG20.V"    "EV.RMG21.V"
##  [77] "EV.RMG26.V"    "BIGOAT"        "EV.RM003.RT"   "EV.RM102.RT"
##  [81] "EV.RM106.RT"   "EV.RM107.RT"   "EV.RM108.RT"   "EV.RM109.RT"
##  [85] "EV.RM111.RT"   "EV.RM112.RT"   "EV.RM114.RT"   "EV.RM116.RT"
##  [89] "EV.RM118.RT"   "EV.RM119.RT"   "EV.RM120.RT"   "EV.RM121.RT"
##  [93] "EV.RM122.RT"   "EV.RM200.RT"   "EV.RM202.RT"   "EV.RM203.RT"
##  [97] "EV.RM204.RT"   "EV.RM205.RT"   "EV.RM206.RT"   "EV.RM207.RT"
## [101] "EV.RM208.RT"   "EV.RM209.RT"   "EV.RM211.RT"   "EV.RM212.RT"
## [105] "EV.RM213.RT"   "EV.RM214.RT"   "EV.RM215.RT"   "EV.RM216.RT"
## [109] "EV.RM219.RT"   "EV.RM300.RT"   "EV.RM302.RT"   "EV.RM303.RT"
## [113] "EV.RM304.RT"   "EV.RM305.RT"   "EV.RM306.RT"   "EV.RM307.RT"
## [117] "EV.RM308.RT"   "EV.RM309.RT"   "EV.RM311.RT"   "EV.RM312.RT"
## [121] "EV.RM313.RT"   "EV.RM314.RT"   "EV.RM315.RT"   "EV.RM316.RT"
## [125] "EV.RM400.RT"   "EV.RM401.RT"   "EV.RM402.RT"   "EV.RM403.RT"
## [129] "EV.RM404.RT"   "EV.RM405.RT"   "EV.RM406.RT"   "EV.RM407.RT"
## [133] "EV.RM411.RT"   "EV.RM412.RT"   "EV.RM413.RT"   "EV.RM414.RT"
## [137] "EV.RM415.RT"   "EV.RMB4.RT"    "EV.RMG04.RT"   "EV.RMG05.RT"
```

```
## [141] "EV.RMG06.RT"   "EV.RMG07.RT"   "EV.RMG08.RT"   "EV.RMG09.RT"
## [145] "EV.RMG10.RT"   "EV.RMG11.RT"   "EV.RMG14.RT"   "EV.RMG16.RT"
## [149] "EV.RMG17.RT"   "EV.RMG18.RT"   "EV.RMG19.RT"   "EV.RMG20.RT"
## [153] "EV.RMG21.RT"   "EV.RMG26.RT"   "EV.HX2.ENA"    "EV.HX2.HWLP1"
## [157] "EV.HX2.HWRT"   "EV.HX2.HWST"   "EV.HX2.HWSTSP" "EV.HX2.V1"
## [161] "EV.HX2.V1P"    "EV.HX2.V2"     "EV.HX2.V2P"    "OATEMP"
```

Get necessary library

```
library(rpart)
```

Split data smaller

```
slice <- subset(evans, select=c(EV.RMG07.V, EV.RMG06.V))
#slice <- subset(evans, select=c(EV.HX2.HWST, OATEMP))
```

Bin data

```
library(mltools)
slice$EV.RMG07.V <- bin_data(slice$EV.RMG07.V, bins=5, binType = "quantile")
slice$EV.RMG06.V <- bin_data(slice$EV.RMG06.V, bins=5, binType = "quantile")

#slice$EV.RMG07.V <- cut(slice$EV.RMG07.V,breaks = 5)
#slice$EV.RMG06.V <- cut(slice$EV.RMG06.V,breaks = 5)
table(slice$EV.RMG07.V)
```

```
##
##   [13.27, 35.81)  [35.81, 40.958) [40.958, 46.286) [46.286, 57.104)
##              137              138              137              137
##    [57.104, 100]
##              138
```

Create tree

```
test_tree = rpart(slice$EV.RMG07.V ~ slice$EV.RMG06.V, method = "class", data=slice)
summary(test_tree)
```

```
## Call:
## rpart(formula = slice$EV.RMG07.V ~ slice$EV.RMG06.V, data = slice,
##     method = "class")
##   n= 687
##
##          CP nsplit rel error    xerror       xstd
## 1 0.1111111      0 1.0000000 1.0728597 0.01669629
## 2 0.0100000      1 0.8888889 0.9198543 0.02106836
##
## Variable importance
## slice$EV.RMG06.V
##              100
```

```
##
## Node number 1: 687 observations,    complexity param=0.1111111
##   predicted class=[35.81, 40.958)  expected loss=0.7991266  P(node) =1
##     class counts:   137   138   137   137   138
##    probabilities: 0.199 0.201 0.199 0.199 0.201
##   left son=2 (138 obs) right son=3 (549 obs)
##   Primary splits:
##       slice$EV.RMG06.V splits as  LR, improve=25.56871, (0 missing)
##
## Node number 2: 138 observations
##   predicted class=[35.81, 40.958)  expected loss=0.5289855  P(node) =0.2008734
##     class counts:    44    65    21     4     4
##    probabilities: 0.319 0.471 0.152 0.029 0.029
##
## Node number 3: 549 observations
##   predicted class=[57.104, 100]    expected loss=0.7559199  P(node) =0.7991266
##     class counts:    93    73   116   133   134
##    probabilities: 0.169 0.133 0.211 0.242 0.244
```

```
plot(test_tree, uniform=TRUE,
    main="Classification of temperature")
text(test_tree, use.n=TRUE, all=TRUE, cex=.8)
```