The University of British Columbia

Irving K. Barber School of Arts and Sciences

DATA 101

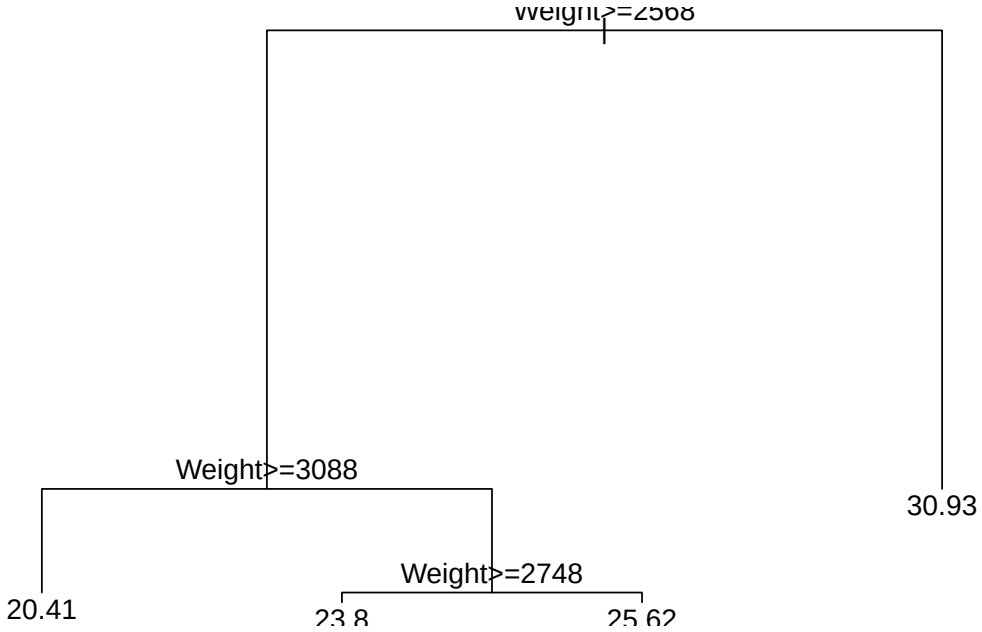
Practice for Advanced Prediction Solutions

1. The data frame `car.test.frame` contains a number of measurements on several cars, including their `Weight` and their gas `Mileage`. Construct a regression tree that relates `Mileage` to `Weight`, using the command

```
library(rpart)
car.tree <- rpart(Mileage ~ Weight, data = car.test.frame)
```

Plot the tree and use the tree to predict the gas mileage for a car with a weight of 2800.

```
plot(car.tree)
text(car.tree)
```



Since 2800 is greater than 2748, and less than 3088, the mileage prediction is 25.62.

2. • Forest fire managers have the following operational task: each day during the fire season, they subjectively predict the total number of lightning- plus human-caused fires in a region for the next week, the next month, and the next year. If they anticipate a large number of fires in a region, they will transfer equipment, airplanes,

and work crews to that region to be ready to fight the fires. This kind of operation is expensive, so mistaken predictions are costly. For short term prediction, weather forecasts and subjective judgement are used to predict fires from the weather conditions but models could provide helpful starting points. For longer term prediction, the fire weather information is of course more uncertain. In particular, the moisture in fine fuels and wind speed can change rapidly.

- In this problem, we will set up rules to determine the probability of a large number of fires (called a *fire flap*) in the next week in the region near Dryden, Ontario, given the current week's data. To set up the rules, use a classification tree based on observational data on weather that has been collected at a weather station in the region near Dryden.[1]

  (a) Read the data from the file *DRYDENflap.R* using the `source()` function. The data consist of 587 weekly observations for the period from 1976 through 1998 in the Dryden district of Northwestern Ontario.[2] The variables in the data frame are as follows:

```
 "NfiresDRY"  -- number of fires this week
 "NfiresKEN"  -- number of fires in Kenora region this week
 "NfiresRED"  -- number of fires in Red Lake region this week
 "Temp"       -- most recent noon temperature
 "RH"         -- most recent relative humidity reading
 "WindDir"    -- most recent wind direction
 "WindSpeed"  -- most recent wind speed
 "Rain"       -- most recent 24 hour rain amount
 "FFMC"       -- a measure of surface level moisture
 "DMC"        -- a measure of intermediate (Duff) level moisture
 "DC"         -- a measure of deep level moisture
 "ISI"        -- a measure of how fast a fire would spread
 "BUI"        -- a measure of how built up the fuel is
 "FWI"        -- a measure of overall fire danger
 "DSR"        -- drought severity rating
"Nextwkflap"  -- an indication of whether the next week
                 had more than 5 fires (Dryden)
```

  Fit the tree to the data in (`DRYDENflap`) using the `rpart()` function in the `rpart` package.

  (b) Plot the tree.

  (c) Use the tree to make predictions as to whether there would be a fire flap next week if the current week had the following characteristics:

  i.
```
##      NfiresDRY NfiresKEN NfiresRED Temp RH
## 1141         0         0         0   10 47
##      WindDir WindSpeed Rain FFMC DMC DC ISI BUI
## 1141     112         8    0 87.7  10 25 4.6  10
##      FWI  DSR Nextwkflap
## 1141   5 0.47          0
```

[1] The data is copyright 1998, Queen's Printer for Ontario, Canada, and was referenced under agreement with the Ontario Ministry of Natural Resources.

[2] The weather observations are from the 10400 station.

(This corresponds to the 260th observation in `DRYDENflap`.)

ii.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 634         3         1         6   26 44     270
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 634         9  0.2   87  17 126 4.4  25   8 108
##     Nextwkflap
## 634          1
```

(This corresponds to the first observation in `DRYDENflap`.)

iii.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 895         0         0         0   21 70     157
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 895         8    0   86  21 267 3.7  34   8 108
##     Nextwkflap
## 895          0
```

(This corresponds to the 138th observation in `DRYDENflap`.)

iv.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 478         6        11         6   18 51       0
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 478         0  0.2   86  20 181 2.5  31   5  47
##     Nextwkflap
## 478          0
```

(This corresponds to the 510th observation in `DRYDENflap`.)

Solution to Question 2:

(a) Read the data from the file *DRYDENflap.R* using the `source()` function.

```
source("DRYDENflap.R")
```

The data consist of 587 weekly observations for the period from 1976 through 1998 in the Dryden district of Northwestern Ontario.[3] The variables in the data frame are as follows:

```
"NfiresDRY"  -- number of fires this week
"NfiresKEN"  -- number of fires in Kenora region this week
"NfiresRED"  -- number of fires in Red Lake region this week
"Temp"       -- most recent noon temperature
"RH"         -- most recent relative humidity reading
"WindDir"    -- most recent wind direction
"WindSpeed"  -- most recent wind speed
"Rain"       -- most recent 24 hour rain amount
"FFMC"       -- a measure of surface level moisture
"DMC"        -- a measure of intermediate (Duff) level moisture
"DC"         -- a measure of deep level moisture
"ISI"        -- a measure of how fast a fire would spread
```

[3]The weather observations are from the 10400 station.

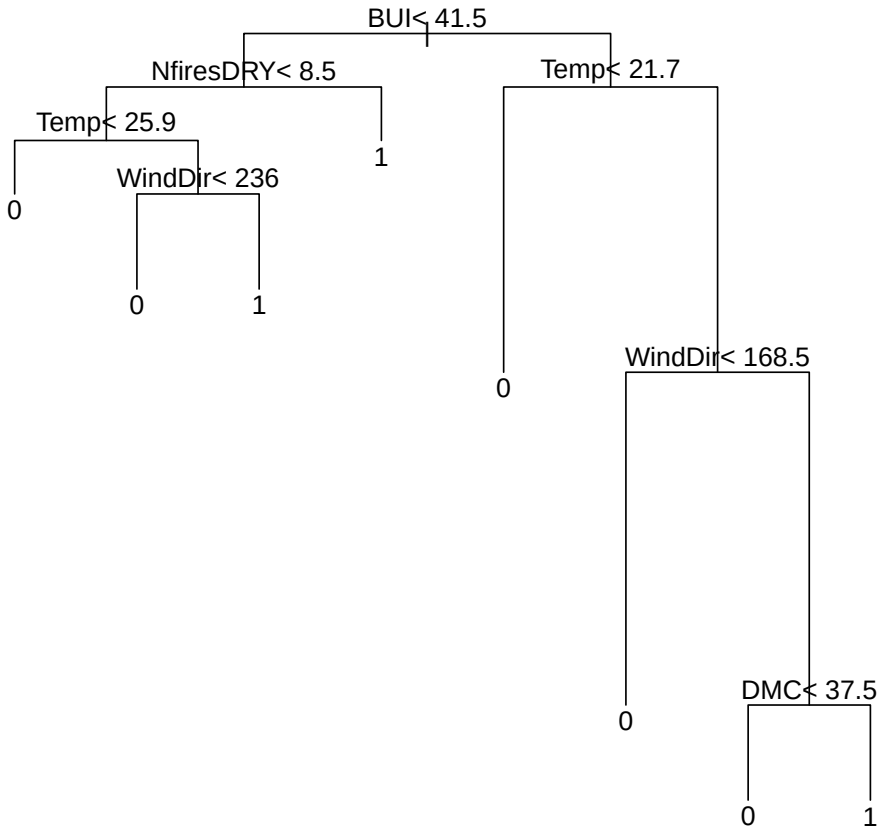
```
 "BUI"        -- a measure of how built up the fuel is
 "FWI"        -- a measure of overall fire danger
 "DSR"        -- drought severity rating
"Nextwkflap"  -- an indication of whether the next week
                 had more than 5 fires (Dryden)
```

Fit the tree to the data in (`DRYDENflap`) using the `rpart()` function in the `rpart` package.

```
library(rpart)  # load package to construct trees
DRY.rpart <- rpart(formula = Nextwkflap ~ ., data = DRYDENflap)
```

(b) Plot the tree.

```
plot(DRY.rpart)
text(DRY.rpart)
```



(c) Use the tree to make predictions as to whether there would be a fire flap next week if the current week had the following characteristics:

i.
```
##      NfiresDRY NfiresKEN NfiresRED Temp RH
## 1141         0         0         0   10 47
##      WindDir WindSpeed Rain FFMC DMC DC ISI BUI
## 1141     112         8    0 87.7  10 25 4.6  10
##      FWI  DSR Nextwkflap
## 1141   5 0.47          0
```

(This corresponds to the 260th observation in `DRYDENflap`.)

```
predict(DRY.rpart, newdata = DRYDENflap[260,], type="class")
## 1141
##    0
## Levels: 0 1
```

A fire flap is not predicted for next week.

ii.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 634         3         1         6   26 44     270
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 634         9  0.2   87  17 126 4.4  25   8 108
##     Nextwkflap
## 634          1
```

(This corresponds to the first observation in `DRYDENflap`.)

```
predict(DRY.rpart, newdata = DRYDENflap[1,], type="class")
## 634
##   1
## Levels: 0 1
```

A fire flap is predicted for next week.

iii.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 895         0         0         0   21 70     157
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 895         8    0   86  21 267 3.7  34   8 108
##     Nextwkflap
## 895          0
```

(This corresponds to the 138th observation in `DRYDENflap`.)

```
predict(DRY.rpart, newdata = DRYDENflap[138,], type="class")
## 895
##   0
## Levels: 0 1
```

A fire flap is not predicted for next week.

iv.
```
##     NfiresDRY NfiresKEN NfiresRED Temp RH WindDir
## 478         6        11         6   18 51       0
##     WindSpeed Rain FFMC DMC  DC ISI BUI FWI DSR
## 478         0  0.2   86  20 181 2.5  31   5  47
##     Nextwkflap
## 478          0
```

(This corresponds to the 510th observation in `DRYDENflap`.)

```
predict(DRY.rpart, newdata = DRYDENflap[510,], type="class")
## 478
##   0
## Levels: 0 1
```

A fire flap is not predicted for next week.