ETC3250 Lab 10

Di Cook

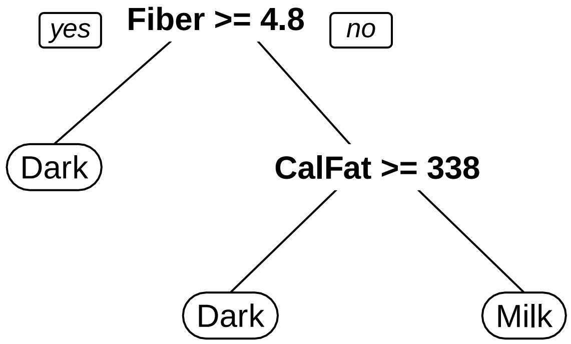
6 October 2015

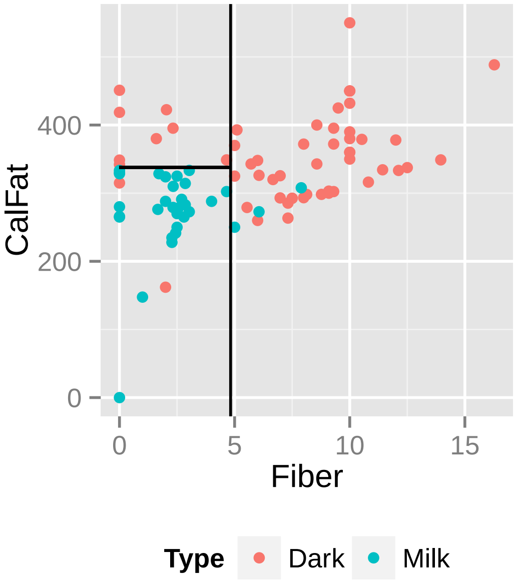
Trees and Forests

Task 1

Read in the chocolates data, from the class web site. Fit a default tree to the tennis data. Print the tree, write the decision rule, compute the error, and make a plot that shows the boundary.

```
##
##        Dark Milk
##   Dark   53    2
##   Milk    3   29
```





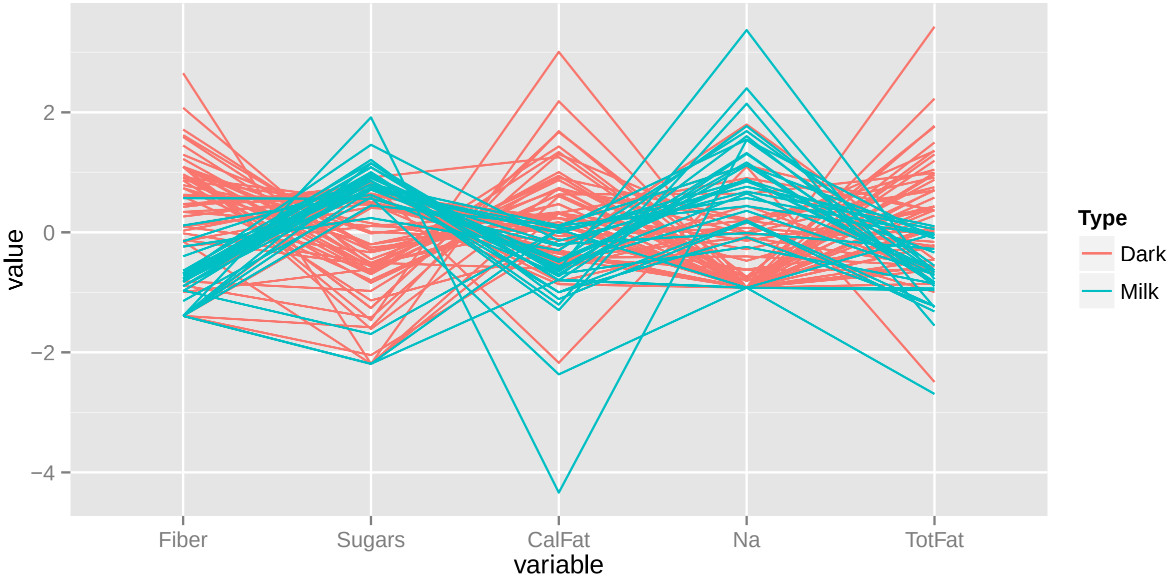
The rule is "Assign to Milk if Fiber is less than 4.83, and Calories from Fat is less than 337.7, otherwise assign to Dark."

Task 2

Fit a random forest to the chocolates data. Report the error, and use a parallel coordinate plot to display the data using the importance to order the variables.

```
##
## Call:
##  randomForest(formula = Type ~ ., data = choc.sub, importance = TRUE,      ntree = 500, mtry = 4)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##         OOB estimate of  error rate: 12.64%
## Confusion matrix:
##      Dark Milk class.error
## Dark   51    4  0.07272727
## Milk    7   25  0.21875000
```

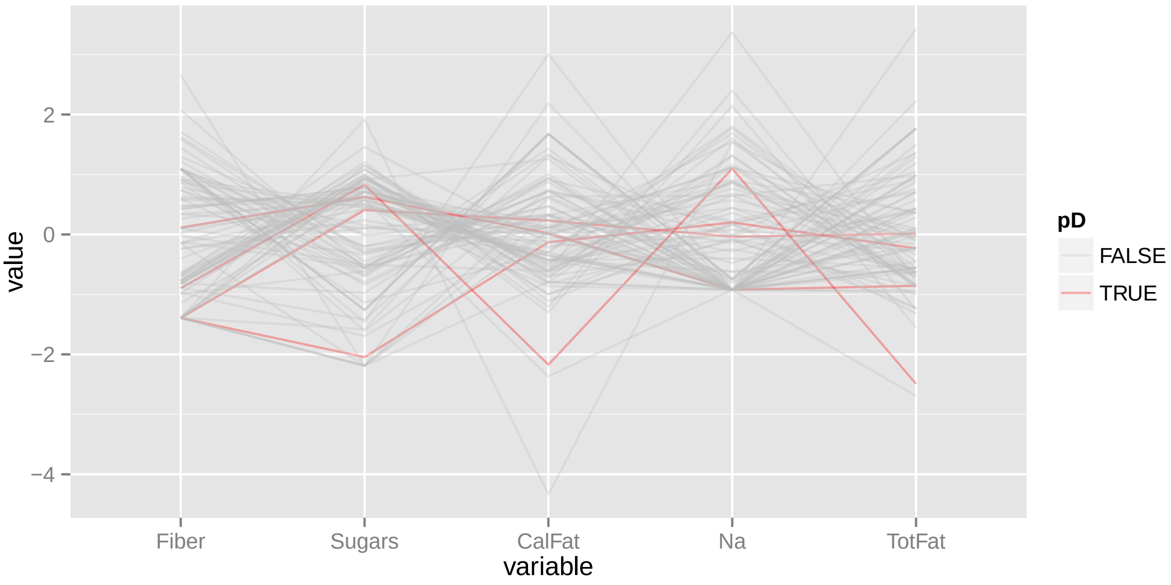
```
##                 Dark        Milk MeanDecreaseAccuracy MeanDecreaseGini
## Fiber     0.052847524 0.141783861          0.083452656        10.019796
## Sugars    0.034996517 0.049000584          0.039583532         6.945043
## CalFat    0.030212919 0.075000832          0.045805667         5.381711
## Na        0.026170750 0.060430313          0.038271930         5.070851
## TotFat    0.032482807 0.050236318          0.038517587         4.601778
## Chol      0.008351639 0.012182225          0.009661550         2.414386
## Carbs     0.005338129 0.010662808          0.006912175         1.692141
## SatFat   -0.001141382 0.012784531          0.003721691         1.370797
## Calories  0.001283562 0.004527897          0.002343168         1.354966
## Protein   0.005274399 0.002581641          0.004308650         1.227398
```



Task 3

Find the labels of the dark chocolates that were misclassified by the forest. Are they the same for both classifiers? Explain why these were misclassified. For example, are they dark chocolates with unusually low fiber?

```
##                                                        Type Calories
## Merci Dark Chocolate France                           Dark 578.9474
## Lindt Dark Chocolate Bar Switzerland                  Dark 400.0000
## Toblerone Dark w/ Honey and Almond Nougat Switzerland Dark 484.8485
## Mars Dark Chocolate Bar US                            Dark 460.0000
```



```
##    Type       Calories         CalFat          TotFat          SatFat    
##  Dark:55   Min.   :400.0   Min.   :162.0   Min.   :18.00   Min.   : 0.00  
##  Milk: 0   1st Qu.:511.6   1st Qu.:309.0   1st Qu.:35.00   1st Qu.:20.47  
##            Median :558.1   Median :348.6   Median :39.53   Median :22.86  
##            Mean   :550.7   Mean   :354.4   Mean   :39.81   Mean   :22.50  
##            3rd Qu.:579.5   3rd Qu.:391.4   3rd Qu.:44.19   3rd Qu.:27.29  
##            Max.   :744.2   Max.   :550.0   Max.   :62.50   Max.   :35.00  
##       Chol              Na              Carbs           Fiber       
##  Min.   : 0.000   Min.   :  0.00   Min.   : 4.00   Min.   : 0.000  
##  1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.:36.60   1st Qu.: 5.627  
##  Median : 0.000   Median :  5.00   Median :48.48   Median : 8.130  
##  Mean   : 4.557   Mean   : 20.53   Mean   :46.27   Mean   : 7.501  
##  3rd Qu.:10.000   3rd Qu.: 38.49   3rd Qu.:56.79   3rd Qu.:10.000  
##  Max.   :34.884   Max.   :121.21   Max.   :74.42   Max.   :16.279  
##      Sugars        Protein      
##  Min.   : 0.00   Min.   : 4.000  
##  1st Qu.:23.63   1st Qu.: 5.857  
##  Median :32.00   Median : 7.500  
##  Mean   :31.47   Mean   : 7.494  
##  3rd Qu.:45.73   3rd Qu.: 9.195  
##  Max.   :53.49   Max.   :11.628  
```

```
##    Type       Calories         CalFat          TotFat          SatFat    
##  Dark: 0   Min.   :270.0   Min.   :  0.0   Min.   :16.50   Min.   : 0.00  
##  Milk:32   1st Qu.:513.6   1st Qu.:265.0   1st Qu.:29.93   1st Qu.:17.49  
##            Median :531.0   Median :279.5   Median :31.41   Median :19.06  
##            Mean   :527.0   Mean   :273.8   Mean   :31.47   Mean   :18.33  
##            3rd Qu.:550.7   3rd Qu.:311.1   3rd Qu.:35.00   3rd Qu.:21.21  
##            Max.   :628.6   Max.   :333.3   Max.   :37.50   Max.   :24.24  
##       Chol             Na             Carbs           Fiber      
##  Min.   : 0.00   Min.   :  0.00   Min.   :25.00   Min.   :0.000  
##  1st Qu.:10.06   1st Qu.: 49.70   1st Qu.:54.41   1st Qu.:1.500  
##  Median :14.29   Median : 79.47   Median :57.95   Median :2.386  
##  Mean   :14.59   Mean   : 76.45   Mean   :57.26   Mean   :2.343  
##  3rd Qu.:23.39   3rd Qu.:102.44   3rd Qu.:60.61   3rd Qu.:2.857  
##  Max.   :30.30   Max.   :191.33   Max.   :70.73   Max.   :7.895  
##      Sugars         Protein     
##  Min.   : 0.00   Min.   : 4.500  
##  1st Qu.:49.62   1st Qu.: 5.929  
##  Median :52.50   Median : 6.857  
##  Mean   :48.48   Mean   : 6.705  
##  3rd Qu.:54.75   3rd Qu.: 7.538  
##  Max.   :70.73   Max.   :10.000  
```

```
##                                                       Calories CalFat
## Merci Dark Chocolate France                                579    342
## Lindt Dark Chocolate Bar Switzerland                       400    315
## Toblerone Dark w/ Honey and Almond Nougat Switzerland      485    326
## Mars Dark Chocolate Bar US                                 460    162
##                                                       TotFat SatFat Chol
## Merci Dark Chocolate France                               37     21   26
## Lindt Dark Chocolate Bar Switzerland                      35     20    0
## Toblerone Dark w/ Honey and Almond Nougat Switzerland     30     18   15
## Mars Dark Chocolate Bar US                                18     12   10
##                                                       Na Carbs Fiber
## Merci Dark Chocolate France                           39    53   0.0
## Lindt Dark Chocolate Bar Switzerland                  50    55   0.0
## Toblerone Dark w/ Honey and Almond Nougat Switzerland  0    61   6.1
## Mars Dark Chocolate Bar US                            90    72   2.0
##                                                       Sugars Protein
## Merci Dark Chocolate France                             44.7     7.9
## Lindt Dark Chocolate Bar Switzerland                     2.5     7.5
## Toblerone Dark w/ Honey and Almond Nougat Switzerland   48.5     6.1
## Mars Dark Chocolate Bar US                              52.0     4.0
```

Merci and Lindt both have 0 Fiber. The others have low calories from fat. Mars Dark Chocolate bas has low Fiber and low calories from fat, low fat, and are at the low end of Sugars for a dark chocolate. It really looks like a milk chocolate.

Task 4

There are a number of zeros in the data. Do you think these are really zeros? How might you fix this?

There are too many zeros. It makes us suspicious that when the information was missing on the nutrition label that the data curators substituted a 0, at least for some of the values. A few of the zeros are believable. We would suggest substituting in a mean or median value for the type of chocolate would help make the classification cleaner.

Assignment

Using the best model that you, tree, forest, lda, svm, … predict the type of chocolate of the `chocolates-new.csv` data provided on the web.

The true labels of the 15 new chocolates are

```
                                 Name            MFR   Country Type
1                          Royal Dark        Cadbury        UK Dark
2                 Rich Dark Chocolate    Darrell Lea Australia Dark
3                 Fine Milk Chocolate    Darrell Lea Australia Milk
4               Divine Milk Chocolate         Divine       USA Milk
5                  85% Dark Chocolate         Divine       USA Dark
6                  365 Organic Swiss     Whole Foods       USA Milk
7         365 Organic Swiss 52% Dark     Whole Foods       USA Dark
8               Dagoba Milk Chocolate        Hershey       USA Milk
9         Organic Very Dark Chocolate Equal Exchange       USA Dark
10             Organic Milk Chocolate Equal Exchange       USA Milk
11 Organic Panama Extra Dark Chocolate Equal Exchange       USA Dark
12                           Chocozoo           Amul     India Milk
13                             Fundoo           Amul     India Milk
14     Chocolaterie Bernard Callebaut      Callebaut    France Milk
15    Callebaut Bittersweet Chocolate      Callebaut    France Dark
```