ETC3250 Lab 10

Di Cook

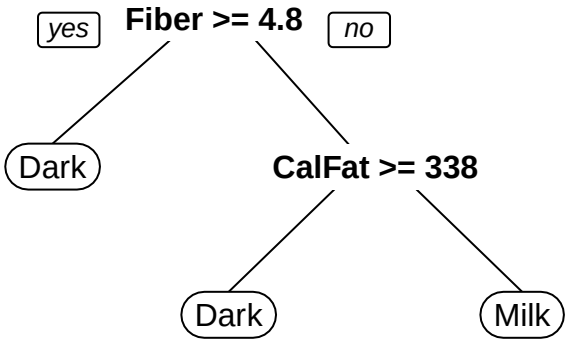
6 October 2015

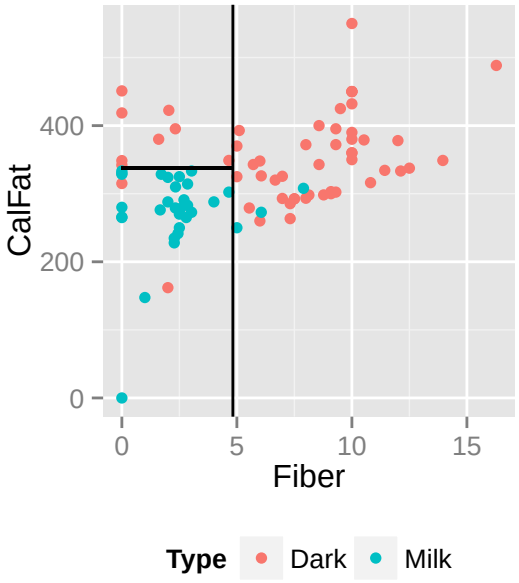
Trees and Forests

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
```





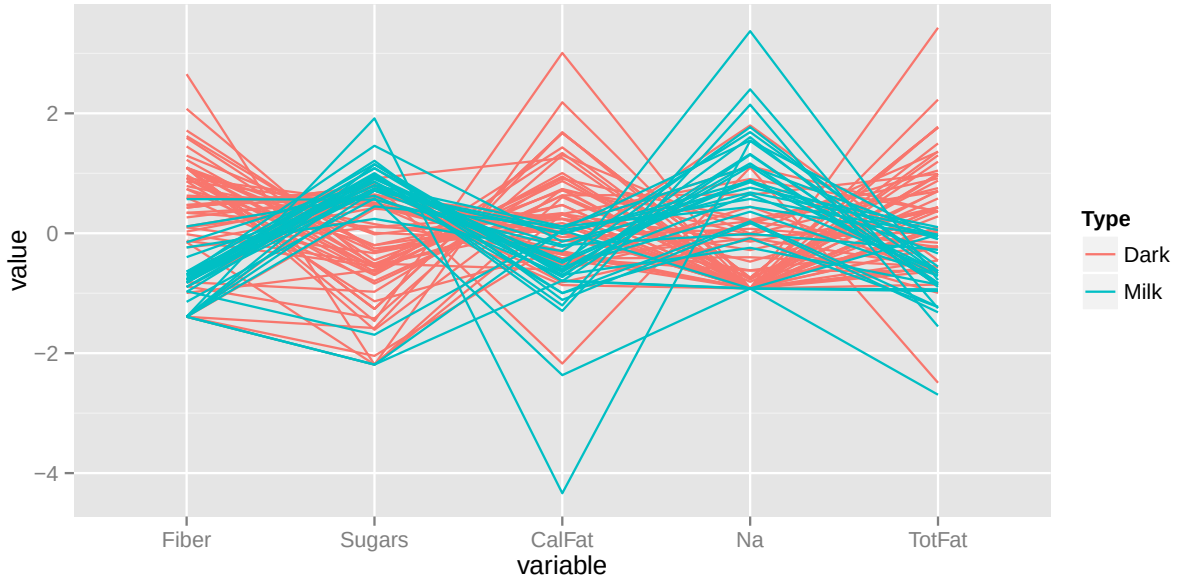
Task 2

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   50    5  0.09090909
## Milk    6   26  0.18750000
```

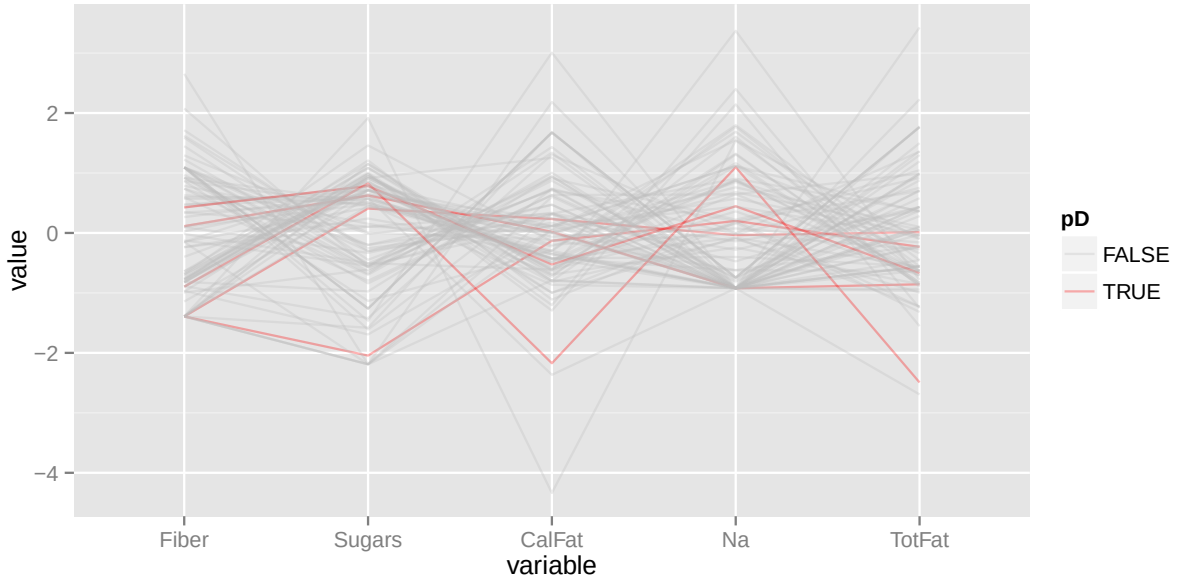
```
##                   Dark         Milk MeanDecreaseAccuracy MeanDecreaseGini
## Calories -0.0010772038  0.003209897          0.0007082395         1.461267
## CalFat    0.0332570540  0.080997356          0.0502527243         5.810786
## TotFat    0.0336386356  0.048383214          0.0389402183         4.813948
## SatFat    0.0007459212  0.016278936          0.0059065295         1.358775
## Chol      0.0033141671  0.014986783          0.0074715959         2.083233
## Na        0.0235610784  0.056902493          0.0361028379         5.135455
## Carbs     0.0026560783  0.005315225          0.0038580361         1.746911
## Fiber     0.0525451279  0.124589862          0.0787407865         9.261681
## Sugars    0.0409808393  0.057056998          0.0463454034         7.101314
## Protein   0.0051522147 -0.002231609          0.0024627721         1.142997
```



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
## Hershey's Rich Dark Chocolate Kisses US            Dark 560.9756
## Mars Dark Chocolate Bar US                         Dark 460.0000
```



Task 4

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

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.