Wk 7: Supervised Learning

Jeanette Henry

1. How do supervised learning algorithms solve regression and classification problems? (not a description of the internal workings of the algorithms)

Supervised learning is capable of making predictions based on a set of data by finding patterns in the data. Classification supervised learning involved prediction a category. Regression supervised learning predicts a value.

1. What packages (in R, Python...) perform supervised learning?
   1. Caret classification and regression training
   2. randomForest
   3. nnett
   4. e1071
   5. kernlab
   6. scikit-learn
   7. pattern
2. How would we compare the results of two different models, or sets of hyperparameters for one model?

Hyperparameters adjust the algorithms or data preparation in predictive models. Crosstable and Carat can be used to evaluate how effective the predicted results are.

head(CO2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Plant** | **Type** | **Treatment** | **conc** | **uptake** |
| Qn1 | Quebec | nonchilled | 95 | 16.0 |
| Qn1 | Quebec | nonchilled | 175 | 30.4 |
| Qn1 | Quebec | nonchilled | 250 | 34.8 |
| Qn1 | Quebec | nonchilled | 350 | 37.2 |
| Qn1 | Quebec | nonchilled | 500 | 35.3 |
| Qn1 | Quebec | nonchilled | 675 | 39.2 |

In [45]:

summary(CO2)

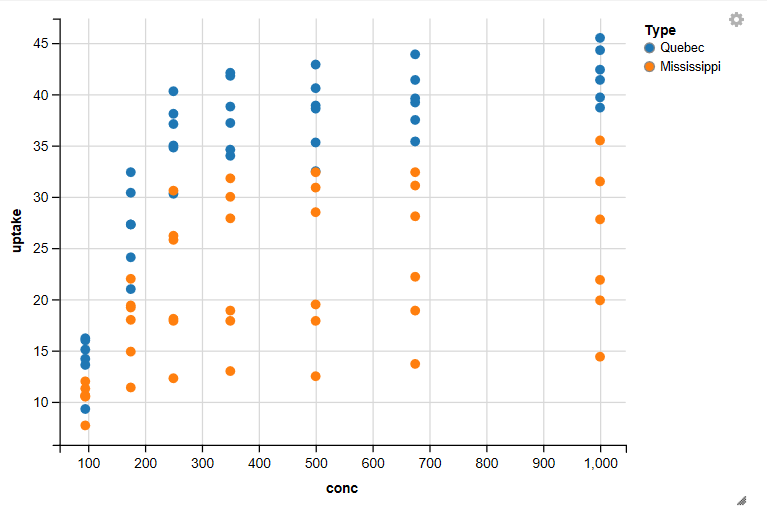
Plant Type Treatment conc uptake   
 Qn1 : 7 Quebec :42 nonchilled:42 Min. : 95 Min. : 7.70   
 Qn2 : 7 Mississippi:42 chilled :42 1st Qu.: 175 1st Qu.:17.90   
 Qn3 : 7 Median : 350 Median :28.30   
 Qc1 : 7 Mean : 435 Mean :27.21   
 Qc3 : 7 3rd Qu.: 675 3rd Qu.:37.12   
 Qc2 : 7 Max. :1000 Max. :45.50   
 (Other):42

In [46]:

library(ggvis)

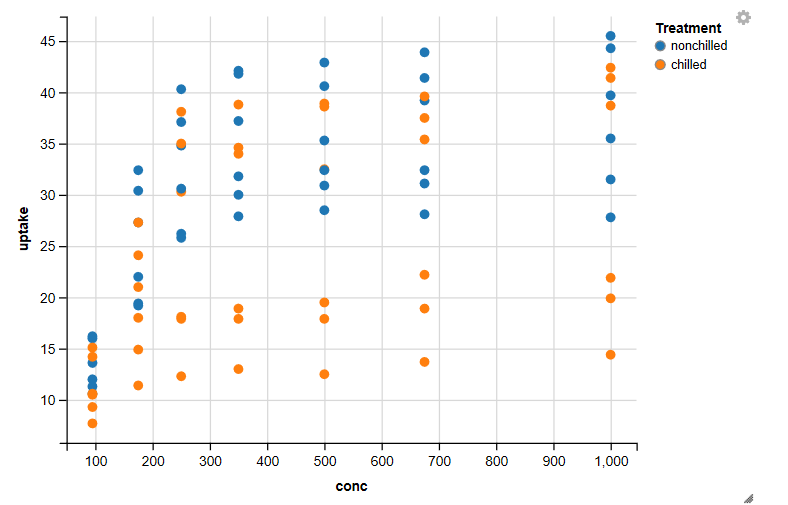
In [50]:

CO2 %>% ggvis(**~**conc, **~**uptake, fill=**~**Type) %>%layer\_points()



In [51]:

CO2 %>% ggvis(**~**conc, **~**uptake, fill=**~**Treatment) %>%layer\_points()



In [ ]:

​

In [52]:

library(class)

In [53]:

set.seed(3465)

In [54]:

ind <- sample(2, nrow(CO2), replace=TRUE, prob=c(0.8, 0.2))

In [57]:

CO2Train <- CO2[ind**==**1, 4**:**5]

CO2Test <- CO2[ind**==**2, 4**:**5]

CO2TrainLabels <- CO2[ind**==**1, 2]

CO2TestLabels <- CO2[ind**==**2, 2]

​

In [58]:

CO2\_pred <- knn(train=CO2Train, test=CO2Test, cl=CO2TrainLabels, k=3)

In [59]:

CO2\_pred

1. Quebec

1. Quebec

1. Quebec

1. Mississippi

1. Quebec

1. Mississippi

1. Quebec

1. Quebec
2. Mississippi

1. Mississippi

1. Quebec

1. Mississippi

1. Mississippi

In [60]:

library(gmodels)

In [61]:

CrossTable(x=CO2TestLabels, y=CO2\_pred, prop.chisq=F, prop.r=F, prop.c=F, prop.t=F)

Cell Contents  
|-------------------------|  
| N |  
|-------------------------|

Total Observations in Table: 13

| CO2\_pred   
CO2TestLabels | Quebec | Mississippi | Row Total |   
--------------|-------------|-------------|-------------|  
 Quebec | 4 | 1 | 5 |   
--------------|-------------|-------------|-------------|  
 Mississippi | 3 | 5 | 8 |   
--------------|-------------|-------------|-------------|  
 Column Total | 7 | 6 | 13 |   
--------------|-------------|-------------|-------------|

In [64]:

library(caret)

In [65]:

set.seed(3256)

In [67]:

trainIndex <- createDataPartition(CO2$Type, p=0.8, list=F, times=1)

In [68]:

head(trainIndex)

|  |
| --- |
| **Resample1** |
| 1 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |

In [69]:

CO2Train <- CO2[trainIndex]

In [70]:

CO2Test <- CO2[**-**trainIndex]

In [71]:

normalize <- **function**(x) {

num <- x **-** min(x)

denom <- max(x) **-** min(x)

return (num**/**denom)

}

​

In [74]:

CO2\_x <- as.data.frame(lapply(CO2[4**:**5], normalize))

In [75]:

summary(CO2)

Plant Type Treatment conc uptake   
 Qn1 : 7 Quebec :42 nonchilled:42 Min. : 95 Min. : 7.70   
 Qn2 : 7 Mississippi:42 chilled :42 1st Qu.: 175 1st Qu.:17.90   
 Qn3 : 7 Median : 350 Median :28.30   
 Qc1 : 7 Mean : 435 Mean :27.21   
 Qc3 : 7 3rd Qu.: 675 3rd Qu.:37.12   
 Qc2 : 7 Max. :1000 Max. :45.50   
 (Other):42

In [76]:

summary(CO2\_x)

conc uptake   
 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.0884 1st Qu.:0.2698   
 Median :0.2818 Median :0.5450   
 Mean :0.3757 Mean :0.5162   
 3rd Qu.:0.6409 3rd Qu.:0.7784   
 Max. :1.0000 Max. :1.0000

In [ ]:

​

In [ ]