1. How do supervised learning algorithms solve regression and classification problems? (not a description of the internal workings of the algorithms)
2. 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
3. How would we compare the results of two different models, or sets of hyperparameters for one model?

> summary(iris)

Sepal.Length Sepal.Width Petal.Length

Min. :4.300 Min. :2.000 Min. :1.000

1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600

Median :5.800 Median :3.000 Median :4.350

Mean :5.843 Mean :3.057 Mean :3.758

3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100

Max. :7.900 Max. :4.400 Max. :6.900

Petal.Width Species

Min. :0.100 setosa :50

1st Qu.:0.300 versicolor:50

Median :1.300 virginica :50

Mean :1.199

3rd Qu.:1.800

Max. :2.500

> summary(iris[c("Petal.Width", "Sepal.Width")])

Petal.Width Sepal.Width

Min. :0.100 Min. :2.000

1st Qu.:0.300 1st Qu.:2.800

Median :1.300 Median :3.000

Mean :1.199 Mean :3.057

3rd Qu.:1.800 3rd Qu.:3.300

Max. :2.500 Max. :4.400

> summary(iris[c("Petal.Length", "Sepal.Length")])

Petal.Length Sepal.Length

Min. :1.000 Min. :4.300

1st Qu.:1.600 1st Qu.:5.100

Median :4.350 Median :5.800

Mean :3.758 Mean :5.843

3rd Qu.:5.100 3rd Qu.:6.400

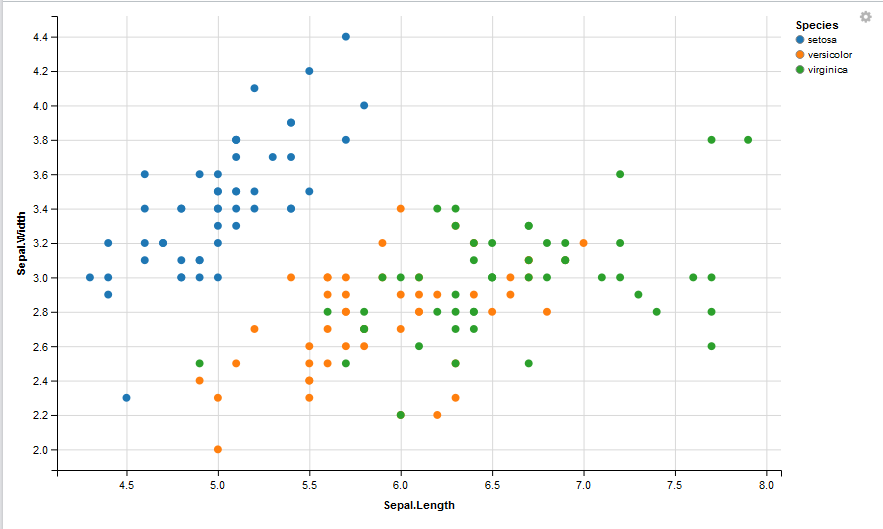
Max. :6.900 Max. :7.900

> library(ggvis)

Warning message:

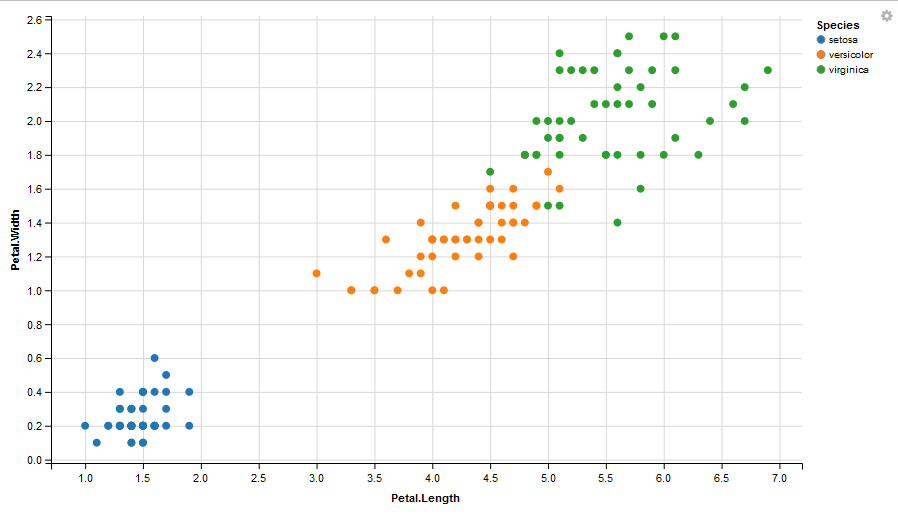
package ‘ggvis’ was built under R version 3.3.3

> iris %>% ggvis(~Sepal.Length, ~Sepal.Width, fill = ~Species) %>% layer\_points()



There is a positive correlation between sepal length and width

|  |
| --- |
| > iris %>% ggvis(~Petal.Length, ~Petal.Width, fill=~Species) %>% layer\_points()  > |
|  |
| |  | | --- | | > | |



There is also a positive correlation for petal length and petal width.

|  |
| --- |
| >  > install.packages("class")  Installing package into ‘C:/Users/Jeane/Documents/R/win-library/3.3’  (as ‘lib’ is unspecified)  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/class\_7.3-14.zip'  Content type 'application/zip' length 101275 bytes (98 KB)  downloaded 98 KB  package ‘class’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\Jeane\AppData\Local\Temp\Rtmps7nY2n\downloaded\_packages  > library(class)  Warning message:  package ‘class’ was built under R version 3.3.3  > set.seed(3465)  > ind <- sample(2, nrow(iris), replace=TRUE, prob=c(0.8, 0.2))  > # use the array, ind, to define the training and test sets  > irisTrain <- iris[ind==1, 1:4]  > irisTest <- iris[ind==2, 1:4]  > irisTrainLabels <- iris[ind==1, 5]  > irisTestLabels <- iris[ind==2, 5]  > source('~/.active-rstudio-document')  > iris\_pred <- knn(train=irisTrain, test=irisTest, cl=irisTrainLabels, k=3)  > iris\_pred  [1] setosa setosa setosa setosa setosa  [6] setosa setosa setosa versicolor versicolor  [11] versicolor versicolor versicolor versicolor versicolor  [16] versicolor virginica virginica virginica virginica  Levels: setosa versicolor virginica  > install.packages("gmodels")  package ‘gtools’ successfully unpacked and MD5 sums checked  package ‘gdata’ successfully unpacked and MD5 sums checked  package ‘gmodels’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\Jeane\AppData\Local\Temp\Rtmps7nY2n\downloaded\_packages  > library(gmodels)  Warning message:  package ‘gmodels’ was built under R version 3.3.3  > CrossTable(x=irisTestLabels, y=iris\_pred, prop.chisq=F, prop.r=F, prop.c=F, prop.t=F)    Cell Contents  |-------------------------|  | N |  |-------------------------|    Total Observations in Table: 20    | iris\_pred  irisTestLabels | setosa | versicolor | virginica | Row Total |  ---------------|------------|------------|------------|------------|  setosa | 8 | 0 | 0 | 8 |  ---------------|------------|------------|------------|------------|  versicolor | 0 | 8 | 0 | 8 |  ---------------|------------|------------|------------|------------|  virginica | 0 | 0 | 4 | 4 |  ---------------|------------|------------|------------|------------|  Column Total | 8 | 8 | 4 | 20 |  ---------------|------------|------------|------------|------------| |
|  |
| |  | | --- | | > | |

> install.packages('caret')

Installing package into ‘C:/Users/Jeane/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

also installing the dependencies ‘iterators’, ‘foreach’, ‘ModelMetrics’

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/iterators\_1.0.8.zip'

Content type 'application/zip' length 316355 bytes (308 KB)

downloaded 308 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/foreach\_1.4.3.zip'

Content type 'application/zip' length 389060 bytes (379 KB)

downloaded 379 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/ModelMetrics\_1.1.0.zip'

Content type 'application/zip' length 560048 bytes (546 KB)

downloaded 546 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/caret\_6.0-76.zip'

Content type 'application/zip' length 2765084 bytes (2.6 MB)

downloaded 2.6 MB

package ‘iterators’ successfully unpacked and MD5 sums checked

package ‘foreach’ successfully unpacked and MD5 sums checked

package ‘ModelMetrics’ successfully unpacked and MD5 sums checked

package ‘caret’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Jeane\AppData\Local\Temp\Rtmps7nY2n\downloaded\_packages

> library(caret)

Loading required package: lattice

Loading required package: ggplot2

Attaching package: ‘ggplot2’

The following object is masked from ‘package:ggvis’:

resolution

Warning messages:

1: package ‘caret’ was built under R version 3.3.3

2: package ‘ggplot2’ was built under R version 3.3.3

> set.seed(3256)

> trainIndex <- createDataPartition(iris$Species, p=0.8, list=F, times=1)

> head(trainIndex)

Resample1

[1,] 1

[2,] 2

[3,] 4

[4,] 6

[5,] 7

[6,] 8

> irisTrain <- iris[trainIndex]

> irisTest <- iris[-trainIndex]

> normalize <- function(x) {

+ num <- x - min(x)

+ denom <- max(x) - min(x)

+ return (num/denom)

+ }

> iris\_x <- as.data.frame(lapply(iris[1:4], normalize))

> summary(iris)

Sepal.Length Sepal.Width Petal.Length

Min. :4.300 Min. :2.000 Min. :1.000

1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600

Median :5.800 Median :3.000 Median :4.350

Mean :5.843 Mean :3.057 Mean :3.758

3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100

Max. :7.900 Max. :4.400 Max. :6.900

Petal.Width Species

Min. :0.100 setosa :50

1st Qu.:0.300 versicolor:50

Median :1.300 virginica :50

Mean :1.199

3rd Qu.:1.800

Max. :2.500

> summary(iris\_x)

Sepal.Length Sepal.Width Petal.Length

Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.2222 1st Qu.:0.3333 1st Qu.:0.1017

Median :0.4167 Median :0.4167 Median :0.5678

Mean :0.4287 Mean :0.4406 Mean :0.4675

3rd Qu.:0.5833 3rd Qu.:0.5417 3rd Qu.:0.6949

Max. :1.0000 Max. :1.0000 Max. :1.0000

Petal.Width

Min. :0.00000

1st Qu.:0.08333

Median :0.50000

Mean :0.45806

3rd Qu.:0.70833

Max. :1.00000

> install.packages("RTextTools")