

Week3 Lab

Jiaju Shen

Wednesday, February 18, 2015

Exercise 1

```
#Almost everyone did a good job for this part.  
#There I won't show this part. Contact me for any questions.
```

Exercise 2

```
# read data in  
aba<-read.csv("http://escience.rpi.edu/data/DA/abalone.csv")  
naba<-dim(aba)[1]  
#90% to train  
sampling.rate=0.9  
#remainder to test  
num.test.set.labels=naba*(1.-sampling.rate)  
#construct a random set of training indices (training)  
training <-sample(1:naba,sampling.rate*naba, replace=FALSE)  
#build the training set (train)  
train<-subset(aba[training,],select=c("Length","Diameter","Height","Whole.weight","Shucked.weight","Viscer  
#construct the remaining test indices (testing)  
testing<-setdiff(1:naba,training)  
#define the test set  
test<-subset(aba[testing,],select=c("Length","Diameter","Height","Whole.weight","Shucked.weight","Viscer  
#construct labels for another variable (Rings) in the training set  
crings<-aba$Rings[training]  
#construct true labels the other variable in the test set  
true.labels<-aba$Rings[testing]  
#run the classifier, can change k  
library(class)  
classif<-knn(train,test,crings,k=5)  
#view the classifier  
classif
```

```
## [1] 8 11 8 10 4 8 16 10 6 8 10 14 15 11 11 7 7 10 8 16 6 9 6  
## [24] 11 8 10 13 7 6 9 9 6 8 15 10 9 10 11 8 12 16 11 9 8 16 9  
## [47] 13 7 6 8 7 10 9 12 12 9 9 7 13 6 9 7 20 12 15 8 17 14 16  
## [70] 10 17 12 10 13 6 8 7 8 6 9 11 9 11 4 6 9 7 8 7 8 8 8  
## [93] 9 11 12 11 5 6 6 6 8 6 7 8 9 9 9 9 8 10 10 10 9 9 13  
## [116] 16 10 11 11 6 6 6 7 7 7 8 9 8 9 8 10 10 11 8 12 9 10 10  
## [139] 9 7 9 9 8 9 9 8 9 9 10 10 7 12 8 9 9 10 9 9 10 10 9  
## [162] 10 10 9 11 14 11 10 10 8 11 10 6 7 8 8 11 9 8 11 10 9 15 16  
## [185] 10 11 10 11 12 12 5 8 10 7 8 7 9 11 9 12 7 7 10 6 8 10 11  
## [208] 9 11 13 5 9 8 10 12 11 7 7 10 10 11 10 11 7 11 14 9 9 9 6
```

```
## [231] 9 8 8 9 16 17 14 11 11 4 13 9 6 5 8 6 7 8 8 9 8 10 10
## [254] 11 9 18 11 12 8 7 7 8 7 9 12 6 7 5 9 7 10 8 9 9 9 10
## [277] 8 10 12 7 6 7 13 10 9 13 13 8 9 8 13 9 9 9 9 11 11 10 11
## [300] 5 9 16 10 10 8 6 9 10 7 10 10 10 6 14 11 8 6 11 10 5 12 12
## [323] 16 9 10 10 7 8 7 7 8 8 15 7 11 11 10 14 4 18 14 13 17 4 8
## [346] 10 14 10 10 8 5 7 8 11 11 9 10 8 10 10 10 11 8 8 9 9 9 11
## [369] 10 9 10 13 12 20 7 7 10 10 9 9 9 10 9 7 8 8 10 9 9 11 5
## [392] 8 11 8 9 11 7 8 7 11 8 11 5 7 6 6 8 10 10 9 7 9 11 7
## [415] 8 9 9 6
## 27 Levels: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 ... 27
```

```
#looks at attributes
attributes(.Last.value)
```

```
## $names
## [1] "help_type"
```

Exercise 3

```
data("iris")
set.seed(321)
#high iteration count
iris.km <- kmeans(iris[, -5], 3, iter.max = 1000)
tbl <- table(iris[, 5], iris.km$cluster)
tbl
```

```
##
##           1  2  3
## setosa      0  0 50
## versicolor  2 48  0
## virginica  36 14  0
```

```
#If we define 4 clusters
iris.km4 <- kmeans(iris[, -5], 4, iter.max = 1000)
tbl2 <- table(iris[, 5], iris.km4$cluster)
tbl2
```

```
##
##           1  2  3  4
## setosa     50  0  0  0
## versicolor  0 23  0 27
## virginica  0 22 27  1
```