Marcus Crowder
LN-4
CIS-3920

4.1 Starting with the 25 cats and dogs, walk through the steps done in the lecture notes, showing your work. Those values are in the accompanying R file under the name Dogs. Don't copy and paste my commands. Rather type in all commands yourself and think about what the commands are you doing. That will help you learn them because you will likely have to make many corrections to your typing!

1. The first thing I did was get the Dogs.csv from my downloads folder.

```
> CatDog = read.csv("~/Downloads/Dogs.csv")
> library(class)
> Dogs = CatDog
> Dogs
   Index Weight Whisker Class Study
1      1   12.1   3.020   Dog Train
2      2    7.2   1.500   Cat Train
3      3    6.8   2.100   Dog Train
4      4   12.2   3.810   Cat Train
5      5   13.2   2.300   Dog Train
6      6   11.3   1.900   Dog Train
7      7   15.3   3.830   Cat Train
8      8    5.3   1.330   Cat Train
9      9    8.5   2.130   Cat Train
10    10    9.5   2.380   Cat Train
11    11    8.8   2.200   Dog Train
12    12    9.1   1.830   Dog Train
13    13    8.5   2.630   Cat Train
14    14    9.5   2.380   Cat Train
```

Then sliced the data as shown by the professor's Examples.

```
> Dogs[1:5, 2:3]
  Weight Whisker
1   12.1    3.02
2    7.2    1.50
3    6.8    2.10
4   12.2    3.81
5   13.2    2.30
> Dogs[1:15,4]
 [1] Dog Cat Dog Cat Dog Dog Cat Cat Cat Cat I
Levels: Cat Dog
> Dogs[16:25,4]
 [1] Dog Cat Cat Cat Cat Dog Dog Cat Cat Dog
Levels: Cat Dog
> TrainX = Dogs[1:15,2:3]
> TrainX
   Weight Whisker
1    12.1    3.02
2     7.2    1.50
3     6.8    2.10
4    12.2    3.81
5    13.2    2.30
```

Creating Variables for TrainX and TestX
TrainY and TestY
> TrainX = Dogs[1:15,2:3]
> TestX = Dogs[16:25, 2:3]
> TrainY = Dogs[1:15,4]
> TestY = Dogs[16:25,4]

-To match the professors Knn output I used a k = 3 for the knn function.

```
> TrainY = Dogs[1:15,4]
> TrainY
 [1] Dog Cat Dog Cat Dog Dog Cat Cat Cat Cat Dog Dog Cat Cat Dog
Levels: Cat Dog
> TestY = Dogs[16:25,4]
> TestY
 [1] Dog Cat Cat Cat Cat Dog Dog Cat Cat Dog
Levels: Cat Dog
> knn.pred = knn( TrainX, TestX, TrainY, 3)
> knn.pred
 [1] Dog Dog Cat Dog Cat Cat Dog Cat Cat Dog
Levels: Cat Dog
> table(knn.pred, TestY)
        TestY
knn.pred Cat Dog
     Cat   4   1
     Dog   2   3
> table.out = table(knn.pred, TestY)
> table.out[1,1]
[1] 4
> (table.out[1,1]+table.out[2,2])/sum(table.out)
[1] 0.7
```

The next step was to open the Professor's SmarketTatum Data. With these commands.
> SMarketTatum <- read.csv("~/Downloads/SMarketTatum.csv")
Then made variables for Lag1 Lag2 Year, train, etc.
There were some difficulties in inputting the commands but I eventually got it right.

```
> SMarketTatum <- read.csv("~/Downloads/SMarketTatum.csv")
> Lag1 <- SMarketTatum$Lag1
> Lag2 <- SMarketTatum$Lag2
> Year <- SMarketTatum$Year
> train = (Year<2005)
> SMarketTatum.2005 = SMarketTatum[!train,]
> Direction <- SMarketTatum$Direction
> Direction.2005 = Direction[!train]
> train.X = cbind(Lag1, Lag2)[!train]
> train.X = cbind(Lag1, Lag2)[train]
> test.X = cbind(Lag1, Lag2)[!train,]
> train.Direction = Direction[train]
> knn.pred = knn(train.X,Test.X,train.Direction,k=1)
Error in knn(train.X, Test.X, train.Direction, k = 1) :
  object 'Test.X' not found
> knn.pred = knn(train.X,test.X,train.Direction,k=1)
Error in knn(train.X, test.X, train.Direction, k = 1) :
  'train' and 'class' have different lengths
> train.X = cbind(Lag1, Lag2)[train,]
> knn.pred = knn(train.X,test.X,train.Direction,k=1)
> table(knn.pred,Direction.2005)
        Direction.2005
knn.pred Down Up
    Down   43 57
    Up     68 84
> knn.pred = knn(train.X,test.X,train.Direction,k=3)
> table(knn.pred,Direction.2005)
        Direction.2005
knn.pred Down Up
    Down   49 56
    Up     62 85
```

4.2
For this example I used my stock market data for MKC I preprocessed the data in excel. The data contained a range of 2532 dates or data points. I had to take out the first two null values in the range. So my calculations started from the 3 column. Anything Market as LowRisk would have to be below the median range of .730004 for the daily return range. Anything above would be marked as high risk.

I	J	K	L	M
Lag1Range	Lag2Range	Risk		Median Rang
		LowRisk		0.730004
0.59		LowRisk		
0.450001	0.59	LowRisk		
0.459999	0.450001	LowRisk		
0.48	0.459999	LowRisk		
0.709999	0.48	LowRisk		
0.560001	0.709999	LowRisk		
0.59	0.560001	LowRisk		
0.329998	0.59	LowRisk		
0.369999	0.329998	LowRisk		
0.439999	0.369999	HIghRisk		
0.790001	0.439999	LowRisk		
0.71	0.790001	LowRisk		
0.41	0.71	LowRisk		
0.550003	0.41	LowRisk		
0.319999	0.550003	LowRisk		
0.66	0.319999	HIghRisk		
1.739998	0.66	HIghRisk		
1.670002	1.739998	HIghRisk		

I reloaded my MKC market data and renamed it after changing the excel file.

With these commands
- ➔ MKCMarket <- read.csv("~/Downloads/MKC.csv")
- ➔ Lag1 <- MKCMarket$Lag1Range
- ➔ Lag2 <- MKCMarket$Lag2Range
- ➔ Risk = MKCMarket$Risk

```
> MKCMarket <- read.csv("~/Downloads/MKC.csv")
> head(MKCMarket)
     Date  Open  High   Low Close Adj.Close Volume    Range Lag1Range Lag2Range    Risk  X Median.Range
1  9/4/07 35.93 36.09 35.50 35.79  28.54640 618700 0.590000        NA        NA LowRisk NA     0.730004
2  9/5/07 35.52 35.72 35.27 35.29  28.14758 538500 0.450001  0.590000        NA LowRisk NA           NA
3  9/6/07 35.45 35.75 35.29 35.34  28.18746 382100 0.459999  0.450001  0.590000 LowRisk NA           NA
4  9/7/07 35.08 35.41 34.93 34.99  27.90830 537900 0.480000  0.459999  0.450001 LowRisk NA           NA
5 9/10/07 35.22 35.26 34.55 34.79  27.74878 470900 0.709999  0.480000  0.459999 LowRisk NA           NA
6 9/11/07 34.82 35.38 34.82 35.33  28.17948 613100 0.560001  0.709999  0.480000 LowRisk NA           NA
```

I slightly deviated from the example that was given above. I found it a bit too complicated with the year ranges so I did something else. I first created the train.X and test.X data points using 101 plot points instead of using [!train]. While using a k of only 1 my results were surprisingly really good my prediction rate was at 70.029703% which I believe is great for financial data points.

```
> train.X=cbind(Lag1,Lag2)[3:103,]
> test.X=cbind(Lag1,Lag2)[104:204,]
> train.Risk=Risk[3:103]
> test.Risk=Risk[104:204]
> knn.pred = knn(train.X,test.X,train.Risk,k=1)
> table(knn.pred,test.Risk)
          test.Risk
knn.pred   HIghRisk LowRisk
  HIghRisk        9      15
  LowRisk        15      62
> table.out = table(knn.pred,test.Risk)
> table.out[1,1]
[1] 9
> (table.out[1,1]+table.out[2,2])/sum(table.out)
[1] 0.7029703
> sum(table.out)
[1] 101
```