NaiveBayes.R

Shraddha Somani

```r
## Import data
library("quantmod")

## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: TTR

## Version 0.4-0 included new data defaults. See ?getSymbols.

library("lubridate")

## Warning: package 'lubridate' was built under R version 3.3.3

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date

library("e1071")
library(xts)


## Set the start and end date of the data to be considered
start_Date = as.Date("2010-01-01")
end_Date = as.Date("2017-04-18")


## Get market data for all symbols making up the NASDAQ 100 Index
Nasdaq100_Symbols <- c("MAT","NFLX","AAPL","QCOM","MAR")
getSymbols(Nasdaq100_Symbols, from = start_Date, to = end_Date)

##     As of 0.4-0, 'getSymbols' uses env=parent.frame() and
##  auto.assign=TRUE by default.
##
##  This  behavior  will be  phased out in 0.5-0  when the call  will
```

```
##   default to use auto.assign=FALSE. getOption("getSymbols.env") and
##   getOptions("getSymbols.auto.assign") are now checked for alternate
defaults
##
##   This message is shown once per session and may be disabled by setting
##   options("getSymbols.warning4.0"=FALSE). See ?getSymbols for more details.

## [1] "MAT"  "NFLX" "AAPL" "QCOM" "MAR"
```

```
nasdaq100 <- data.frame(as.xts(merge(MAT,NFLX,AAPL,QCOM,MAR)))
head(nasdaq100)
```

```
##            MAT.Open MAT.High MAT.Low MAT.Close MAT.Volume MAT.Adjusted
## 2010-01-04    20.33    20.33   19.96     19.99    2975900     14.72519
## 2010-01-05    19.90    20.06   19.33     19.58    5206100     14.42317
## 2010-01-06    19.50    19.58   19.19     19.50    6354900     14.36424
## 2010-01-07    19.70    20.07   19.64     20.00    9714500     14.73256
## 2010-01-08    19.91    20.06   19.74     19.77    3362600     14.56313
## 2010-01-11    19.79    19.98   19.66     19.90    5180500     14.65889
##            NFLX.Open NFLX.High NFLX.Low NFLX.Close NFLX.Volume
## 2010-01-04     55.52     55.73    52.96      53.48    17239600
## 2010-01-05     53.57     53.60    50.81      51.51    23753100
## 2010-01-06     51.53     53.71    50.38      53.32    23290400
## 2010-01-07     54.12     54.30    52.24      52.40     9955400
## 2010-01-08     52.49     54.20    52.26      53.30     8180900
## 2010-01-11     53.62     53.93    52.70      53.23     6783700
##            NFLX.Adjusted AAPL.Open AAPL.High AAPL.Low AAPL.Close
## 2010-01-04      7.640000    213.43    214.50   212.38     214.01
## 2010-01-05      7.358572    214.60    215.59   213.25     214.38
## 2010-01-06      7.617143    214.38    215.23   210.75     210.97
## 2010-01-07      7.485714    211.75    212.00   209.05     210.58
## 2010-01-08      7.614286    210.30    212.00   209.06     211.98
## 2010-01-11      7.604286    212.80    213.00   208.45     210.11
##            AAPL.Volume AAPL.Adjusted QCOM.Open QCOM.High QCOM.Low
## 2010-01-04   123432400      27.72704     46.79     47.31    46.57
## 2010-01-05   150476200      27.77498     47.52     48.32    47.50
## 2010-01-06   138040000      27.33318     48.26     48.43    47.38
## 2010-01-07   119282800      27.28265     47.75     49.02    47.55
## 2010-01-08   111902700      27.46403     49.11     49.80    48.89
## 2010-01-11   115557400      27.22176     49.58     49.69    48.64
##            QCOM.Close QCOM.Volume QCOM.Adjusted MAR.Open MAR.High MAR.Low
## 2010-01-04      46.94    14570200      39.76652    27.61    27.93   27.20
## 2010-01-05      48.07    24241000      40.72383    27.57    27.92   27.37
## 2010-01-06      47.60    22743800      40.32566    27.74    28.02   27.53
## 2010-01-07      48.98    25266300      41.49476    27.75    28.17   27.70
## 2010-01-08      49.47    25938200      41.90988    27.78    28.17   27.76
## 2010-01-11      49.29    14672400      41.75739    28.13    28.34   27.81
##            MAR.Close MAR.Volume MAR.Adjusted
## 2010-01-04     27.63    2860300     23.77412
## 2010-01-05     27.79    2627800     23.91180
```

```
## 2010-01-06      27.87    2789800     23.98063
## 2010-01-07      27.98    2560700     24.07528
## 2010-01-08      28.04    3961000     24.12691
## 2010-01-11      27.98    2714200     24.07528

## Set outcome variable
outcomeSymbol <- 'AAPL.Close'
nasdaq100 <- xts(nasdaq100,order.by=as.Date(rownames(nasdaq100)))
nasdaq100 <- as.data.frame(merge(nasdaq100,
lm1=lag(nasdaq100[,outcomeSymbol],-1)))
nasdaq100$outcome <- ifelse(nasdaq100[,paste0(outcomeSymbol,'.1')] >
nasdaq100[,outcomeSymbol], 1, 0)
nasdaq100$date <- as.Date(row.names(nasdaq100))
nasdaq100 <- nasdaq100[order(as.Date(nasdaq100$date, "%m/%d/%Y"), decreasing
= TRUE),]

## Delete the last row from our data frame as it doesn't have an outcome
variable (that is in the future): drop most recent entry as we don't have an
outcome
nasdaq100 <- nasdaq100[2:nrow(nasdaq100),]
dput(names(nasdaq100)[grepl('MAT',names(nasdaq100))])

## c("MAT.Open", "MAT.High", "MAT.Low", "MAT.Close", "MAT.Volume",
## "MAT.Adjusted")

## Retrieving Apple's daily OHLCV from Yahoo Finance
nasdaq100$wday  <- as.POSIXlt(nasdaq100$date)$wday


## Find the day of the week
head(nasdaq100)

##            MAT.Open MAT.High MAT.Low MAT.Close MAT.Volume MAT.Adjusted
## 2017-04-17    25.09    25.27   24.95     25.24    3134900        25.24
## 2017-04-13    25.74    25.78   25.01     25.01    4042400        25.01
## 2017-04-12    25.68    25.97   25.62     25.73    4888700        25.73
## 2017-04-11    25.39    25.74   25.20     25.73    2669700        25.73
## 2017-04-10    24.92    25.41   24.92     25.39    3036100        25.39
## 2017-04-07    25.35    25.39   24.88     24.92    3906400        24.92
##            NFLX.Open NFLX.High NFLX.Low NFLX.Close NFLX.Volume
## 2017-04-17    144.43    147.32   144.43     147.25    14648700
## 2017-04-13    144.25    144.55   142.76     142.92     3621300
## 2017-04-12    144.85    145.74   143.55     143.83     4377200
## 2017-04-11    144.28    144.54   141.98     144.35     4613600
## 2017-04-10    143.04    144.65   141.21     143.85     5760500
## 2017-04-07    143.79    143.89   142.71     143.11     3021100
##            NFLX.Adjusted AAPL.Open AAPL.High AAPL.Low AAPL.Close
## 2017-04-17        147.25    141.48    141.88   140.87     141.83
## 2017-04-13        142.92    141.91    142.38   141.05     141.05
## 2017-04-12        143.83    141.60    142.15   141.01     141.80
```

```
## 2017-04-11            144.35     142.94     143.35   140.06       141.63
## 2017-04-10            143.85     143.60     143.88   142.90       143.17
## 2017-04-07            143.11     143.73     144.18   143.27       143.34
##            AAPL.Volume AAPL.Adjusted QCOM.Open QCOM.High QCOM.Low
## 2017-04-17    16424000        141.83     52.99     53.13    52.75
## 2017-04-13    17652900        141.05     53.13     53.68    52.78
## 2017-04-12    20238900        141.80     54.64     54.65    53.12
## 2017-04-11    30275300        141.63     56.32     56.36    54.72
## 2017-04-10    18473000        143.17     56.43     56.81    56.27
## 2017-04-07    16621300        143.34     56.58     56.60    56.19
##            QCOM.Close QCOM.Volume QCOM.Adjusted MAR.Open MAR.High MAR.Low
## 2017-04-17      52.89    10316800         52.89    91.91    91.98   91.60
## 2017-04-13      52.79    16971600         52.79    91.70    92.12   91.32
## 2017-04-12      53.39    24731300         53.39    91.98    92.00   90.88
## 2017-04-11      55.35    17291800         55.35    91.16    91.59   90.79
## 2017-04-10      56.52     5917100         56.52    91.04    91.72   90.66
## 2017-04-07      56.32     5226900         56.32    91.73    91.73   90.25
##            MAR.Close MAR.Volume MAR.Adjusted AAPL.Close.1 outcome
## 2017-04-17     91.95    1284000        91.95       141.20       0
## 2017-04-13     91.55    2240900        91.55       141.83       1
## 2017-04-12     91.75    3416300        91.75       141.05       0
## 2017-04-11     91.40    2730400        91.40       141.80       1
## 2017-04-10     91.06    2126700        91.06       141.63       0
## 2017-04-07     91.21    2954500        91.21       143.17       0
##                  date wday
## 2017-04-17 2017-04-17    1
## 2017-04-13 2017-04-13    4
## 2017-04-12 2017-04-12    3
## 2017-04-11 2017-04-11    2
## 2017-04-10 2017-04-10    1
## 2017-04-07 2017-04-07    5

PriceDiff <- (nasdaq100$NFLX.Close) - (nasdaq100$NFLX.Open)


## Find the difference between the close price and open price
TrendClass <- ifelse(PriceDiff>0,"UP","DOWN")


## Convert to a binary classification. (In our data set, there are no bars
with an exactly 0 price change so, for simplicity sake, we will not address
bars that had the same open and close price.)
D1 <- data.frame(nasdaq100$wday,TrendClass)


## Create our data set
M1 <- naiveBayes(D1[,1],D1[,2])
M1
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = D1[, 1], y = D1[, 2])
##
## A-priori probabilities:
## D1[, 2]
##      DOWN        UP
## 0.5136314 0.4863686
##
## Conditional probabilities:
##          x
## D1[, 2]      [,1]      [,2]
##    DOWN 3.022293 1.409144
##    UP   3.022422 1.388405
```

#Analysis:

```
MA5 <- EMA(nasdaq100$NFLX.Open,n = 5)
MA10 <- EMA(nasdaq100$NFLX.Open,n = 10)
MACross <- MA5 - MA10

MACross <- round(MACross,2)
```

Lets create a new dataset and split it into a training and test set so we are able to see how well our model does over new data

```
D2 <- data.frame(nasdaq100$wday,MACross, TrendClass)
D2 <- D2[-c(1:10),]
# We need to remove the instances where the 10-period moving average is still being calculated
TrainingData <- D2[1:1216,]
TestData <- D2[1217:1823,]
MACrossModel <- naiveBayes(TrainingData[,1:2],TrainingData[,3])
MACrossModel
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = TrainingData[, 1:2], y = TrainingData[,
##      3])
##
## A-priori probabilities:
## TrainingData[, 3]
##      DOWN        UP
## 0.5065789 0.4934211
##
## Conditional probabilities:
##                   nasdaq100.wday
```

```
## TrainingData[, 3]      [,1]      [,2]
##              DOWN 3.030844 1.417896
##              UP   3.013333 1.386733
##
##                  MACross
## TrainingData[, 3]      [,1]      [,2]
##              DOWN  0.1139448 10.83784
##              UP   -0.4459833 12.66507
```

The Conditional Probability of the EMA Cross, a numeric variable, shows the
mean value for each case ([,1]), and the standard deviation ([,2]). We can
see that the mean difference between the 5-period EMA and 10-period EMA for
long and short trades was 0.54 and - 0.24, respectively.

```
table(predict(MACrossModel,TestData),TestData[,3],dnn=list('predicted','actua
l'))
```

```
##          actual
## predicted DOWN  UP
##     DOWN   313 287
##     UP       6   1
```