Module 11

Neha Parulekar

April 4, 2016

# Additional packages needed

* If necessary install the followings packages. install.packages("RCurl");  
  install.packages("plyr");  
  install.packages("forecast");

library(RCurl)

## Loading required package: bitops

library(plyr)

## Warning: package 'plyr' was built under R version 3.2.4

library(forecast)

## Warning: package 'forecast' was built under R version 3.2.4

## Loading required package: zoo

##   
## Attaching package: 'zoo'

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

## Loading required package: timeDate

## This is forecast 6.2

# Assingment

Answer the following questions:  
 \* Can some form of Time series analysis help in your research project to forcast a prediction?  
 \* If it can apply Time series analysis to your research project? Does it help?  
 \* If (and only if) you can't use some form of Time series analysis help in your research project then apply a form of Time series analysis to the data the Twiiter time series data set M11\_Tweets\_Miley\_Nikki\_Taylor.csv (it is online at '<http://nikbearbrown.com/YouTube/MachineLearning/M11/M11_Tweets_Miley_Nikki_Taylor.csv>') . Does it help?  
 \* Note you only need to use ONE forcasting approach from Module 11, so there will be only ONE assingment for all the modules and the same assingment for all the modules.

**My data does not require time series analysis. Thus I am using Twitter data given. From the forecast plot we can predict that the no. of tweets remain the same in the coming year. They do not change significantly.**

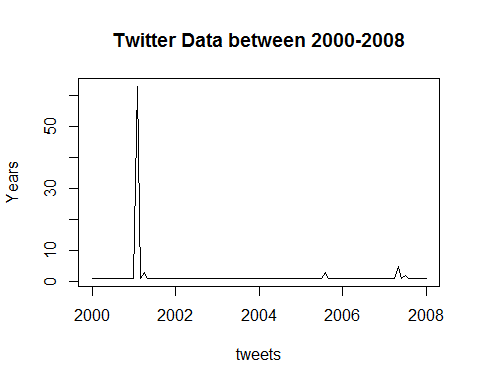
data <- "http://nikbearbrown.com/YouTube/MachineLearning/M11/M11\_Tweets\_Miley\_Nikki\_Taylor.csv"  
data <- read.csv(url(data))  
names(data)

## [1] "X6.60163E.17"   
## [2] "Fri.Oct.30.18.33.50..0000.2015"   
## [3] "https...t.co.V2kb7Qk6VD.Check.out.my..remix.here.for..taylorswift.if.you.didnt.catch.it.already."

head(data ,n = 2)

## X6.60163E.17 Fri.Oct.30.18.33.50..0000.2015  
## 1 6.60163E+17 Fri Oct 30 18:33:51 +0000 2015  
## 2 6.60163E+17 Fri Oct 30 18:33:52 +0000 2015  
## https...t.co.V2kb7Qk6VD.Check.out.my..remix.here.for..taylorswift.if.you.didnt.catch.it.already.  
## 1 Nicki Minaj Strips Down To Bra &amp;amp; Thong For 2015 Calendar!... https://t.co/G5qZeKCTEu https://t.co/8Yz9LRhYbk  
## 2 RT @FactsAboutNM: Nicki Minaj is on Jhene Aiko's New Album featured. Imagine the SLAYAGE https://t.co/KWrhdtNann

TweetCount <- count(data)  
  
# set the freq parameter to 12 to indicate monthly readings  
# ts() function to create a new time series  
Tweet\_timeseries <- ts(TweetCount$freq, start = c(2000, 1), end = c(2008, 1), frequency = 12)   
  
plot(Tweet\_timeseries, xlab='tweets' ,ylab='Years', main='Twitter Data between 2000-2008')



## ------------- USE ARIMA MODEL --------------------  
  
#creating ranges of possible values for the order parameters p, d, and q.  
d <- 0 : 2  
p <- 0 : 6  
q <- 0 : 6  
  
twitter\_models <- expand.grid(d = d, p = p, q = q)  
head(twitter\_models, n = 4)

## d p q  
## 1 0 0 0  
## 2 1 0 0  
## 3 2 0 0  
## 4 0 1 0

getTSModelAIC <- function(ts\_data, p, d, q) {  
  
ts\_model <- arima(ts\_data, order = c(p, d, q))  
 return(ts\_model$aic)  
}  
  
getTSModelAICSafe <- function(ts\_data, p, d, q) { result = tryCatch({  
 getTSModelAIC(ts\_data, p, d, q)  
 },   
error = function(e) {  
 Inf  
 })  
}  
  
# PICK THE BEST MODEL THAT HAS THE SMALLEST AIC   
twitter\_models$aic <- mapply(function(x, y, z)   
getTSModelAICSafe(Tweet\_timeseries, x, y, z), twitter\_models$p,   
 twitter\_models$d, twitter\_models$q)

## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1  
  
## Warning in arima(ts\_data, order = c(p, d, q)): possible convergence  
## problem: optim gave code = 1

subset(twitter\_models,aic == min(aic))

## d p q aic  
## 23 1 0 1 634.5408

# ARIMA model for best p,d,q order model   
twitter\_model <- arima(Tweet\_timeseries, order = c(1, 1, 1))  
summary(twitter\_model)

##   
## Call:  
## arima(x = Tweet\_timeseries, order = c(1, 1, 1))  
##   
## Coefficients:  
## ar1 ma1  
## -0.0076 -0.9843  
## s.e. 0.1046 0.0507  
##   
## sigma^2 estimated as 40.22: log likelihood = -315.27, aic = 636.54  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.5064451 6.30928 1.849631 -114.0283 118.5972 1.250455  
## ACF1  
## Training set -0.01403581

#------------------- Prediction ---------------------------------  
plot(forecast(twitter\_model, 10))

