mt2finalcode

Avery Kan

April 18, 2017

##load packages  
library(DataComputing)  
library(forecast)

DATASET 1  
#1. Read in and explore  
ts1 <- read.csv("q1\_train.csv", as.is = TRUE)  
ts.1 <- ts(ts1[,2])  
ts1 <- sqrt(ts.1+abs(min(ts.1))+0.001)  
plot(ts1, type = 'l')  
#2. Infer from acf and pacf  
ts1.d <- diff(ts1)  
ts1.acf <- ts1.d %>% acf(20)   
ts1.pacf <- ts1.d %>% pacf(20)  
ts1.acf[1]  
ts1.pacf[1]  
ts1.d.s <- ts1.d %>% diff(52)  
ts1.sacf <- ts1.d.s %>% acf(200)  
ts1.psacf <- ts1.d.s %>% pacf(200)  
ts1.sacf[52]   
ts1.psacf[52]  
#3. Write MSE Function  
mse <- function(v1,v2,tsd,n){  
 n1list <- round(seq(from = ceiling(n/2), to = n-1, length.out = 30))  
 test.error <- numeric(length(n1list))  
 for(i in 1:length(n1list)){  
 n1 <- n1list[i]  
 train <- tsd[1:n1]  
 test <- tsd[(n1+1):n]  
 train.arima <- arima(train, order = v1,   
 seasonal = list(order = v2, period = 52))   
 test.pred <- forecast(train.arima,length(test))$mean  
 test.error[i] <- sqrt(sum((test.pred-test)^2))  
 }  
 return(mean(test.error))  
}  
  
#4. Model Testing  
mse(c(3,1,2),c(0,1,1),ts1,length(ts1)) #2.352  
mse(c(2,1,2),c(0,1,1),ts1,length(ts1)) #2.346  
mse(c(1,1,2),c(0,1,1),ts1,length(ts1)) #2.333  
  
mse(c(1,1,1),c(0,1,1),ts1,length(ts1)) #2.331  
mse(c(1,1,1),c(1,1,0),ts1,length(ts1)) #2.092  
  
#5. Predict and plot  
model <- arima(ts1, order = c(1,1,1), seasonal = list(order = c(1,1,0), period = 52))  
  
predict.arima <- predict(model,n.ahead= 104)  
ts1.pred <- (predict.arima$pred)^2 - 0.001 - abs(min(ts.1))  
tsdiag(model)  
  
plot(1:(length(ts.1) + length(ts1.pred)), c(ts.1, ts1.pred), type = 'l', col = 1,  
 main="Predictions", sub="ARIMA(1,1,1)x(1,1,0)\_52",   
 xlab="time", ylab="ts1")  
points((length(ts.1) + 1) : (length(ts1) + length(ts1.pred)), ts1.pred, type = 'l', col = 2)  
  
#6. Write table and verify   
write.table(ts1.pred,  
 sep = ",",  
 col.names = FALSE,  
 row.names = FALSE,  
 file = "Q1\_Avery\_Kan\_25363807.txt")  
  
temp <- read.table("Q1\_Avery\_Kan\_25363807.txt", sep = ",")  
plot(as.numeric(unlist(temp)))  
  
#######################################################################  
DATASET 2

#1. Read in and explore  
  
ts2 <- read.csv("q2\_train.csv", as.is = TRUE)  
ts2 <- ts(ts2[,2])  
  
plot(ts2, type = 'l')

#2. Infer from acf and pacf  
  
acf(ts2)  
ts2.d <- diff(ts2)  
ts2.acf <- ts2.d %>% acf(20)   
ts2.pacf <- ts2.d %>% pacf(20)  
ts2.acf[1]  
ts2.pacf[1]  
ts2.d.s <- ts2.d %>% diff(52)  
ts2.sacf <- ts2.d.s %>% acf(200)  
ts2.psacf <- ts2.d.s %>% pacf(200)  
ts2.sacf[52]   
ts2.psacf[52]  
#3. Model Testing  
  
mse(c(3,1,2),c(0,1,1),ts2,length(ts2)) #9.13  
mse(c(2,1,2),c(0,1,1),ts2,length(ts2)) #9.02  
mse(c(1,1,2),c(0,1,1),ts2,length(ts2)) #8.96  
mse(c(1,1,1),c(0,1,1),ts2,length(ts2)) #9.01  
#4. Predict and plot  
  
model <- arima(ts2, order = c(1,1,2), seasonal = list(order = c(0,1,1), period = 52))  
predict.arima <- predict(model,n.ahead= 104)  
ts2.pred <- predict.arima$pred  
  
plot(1:(length(ts2) + length(ts2.pred)), c(ts2, ts2.pred), type = 'l', col = 1)  
points((length(ts2) + 1) : (length(ts2) + length(ts2.pred)), ts2.pred, type = 'l', col = 2)  
  
#5. Write table and verify   
write.table(ts2.pred,  
 sep = ",",  
 col.names = FALSE,  
 row.names = FALSE,  
 file = "Q2\_Avery\_Kan\_25363807.txt")  
  
temp <- read.table("Q2\_Avery\_Kan\_25363807.txt", sep = ",")  
plot(as.numeric(unlist(temp)))  
  
  
#######################################################################

DATASET 3

#1. Read in and explore  
ts3 <- read.csv("q3\_train.csv", as.is = TRUE)  
ts3 <- ts(ts3[,2])  
plot(ts3, type = 'l')  
#2. Infer from acf and pacf  
acf(ts3)  
ts3.d <- diff(ts3)  
ts3.acf <- ts3.d %>% acf(20)   
ts3.pacf <- ts3.d %>% pacf(20)  
ts3.acf[1]  
ts3.pacf[1]  
ts3.d.s <- ts3.d %>% diff(52)  
ts3.sacf <- ts3.d.s %>% acf(200)  
ts3.psacf <- ts3.d.s %>% pacf(200)  
ts3.sacf[52]   
ts3.psacf[52]  
  
#3. Model testing  
mse(c(3,1,2),c(0,1,1),ts3,length(ts3)) #9.54  
mse(c(2,1,2),c(0,1,1),ts3,length(ts3)) #9.53  
mse(c(1,1,2),c(0,1,1),ts3,length(ts3)) #9.54  
mse(c(1,1,1),c(0,1,1),ts3,length(ts3)) #9.56  
mse(c(1,1,1),c(1,1,0),ts3,length(ts3)) #8.64  
#4. Predict and plot  
model <- arima(ts3, order = c(1,1,1), seasonal = list(order = c(1,1,0), period = 52))  
predict.arima <- predict(model,n.ahead= 104)  
ts3.pred <- predict.arima$pred  
  
plot(1:(length(ts3) + length(ts3.pred)), c(ts3, ts3.pred), type = 'l', col = 1)  
points((length(ts3) + 1) : (length(ts3) + length(ts3.pred)), ts3.pred, type = 'l', col = 2)  
#5. Write table and verify   
write.table(ts3.pred,  
 sep = ",",  
 col.names = FALSE,  
 row.names = FALSE,  
 file = "Q3\_Avery\_Kan\_25363807.txt")  
  
  
temp <- read.table("Q3\_Avery\_Kan\_25363807.txt", sep = ",")  
plot(as.numeric(unlist(temp)))  
  
####################################################################### DATASET 4

#1. Read in and explore  
ts4 <- read.csv("q4\_train.csv", as.is = TRUE)  
ts4 <- ts(ts4[,2])  
#2. Infer from acf and pacf  
acf(ts4)  
ts4.d <- diff(ts4)  
ts4.acf <- ts4.d %>% acf(20)   
ts4.pacf <- ts4.d %>% pacf(20)  
ts4.acf[1]  
ts4.pacf[1]  
ts4.d.s <- ts4.d %>% diff(52)  
ts4.sacf <- ts4.d.s %>% acf(200)  
ts4.psacf <- ts4.d.s %>% pacf(200)  
ts4.sacf[52]   
ts4.psacf[52]  
#3. Model testing  
mse(c(2,1,1),c(0,1,1),ts4,length(ts4)) #6.58  
mse(c(1,1,1),c(0,1,1),ts4,length(ts4)) #6.74  
mse(c(1,1,1),c(1,1,0),ts4,length(ts4)) #error  
#4. Predict and plot  
model <- arima(ts4, order = c(1,1,1), seasonal = list(order = c(0,1,1), period = 52))  
predict.arima <- predict(model,n.ahead= 104)  
ts4.pred <- predict.arima$pred  
  
plot(1:(length(ts4) + length(ts4.pred)), c(ts4, ts4.pred), type = 'l', col = 1)  
points((length(ts4) + 1) : (length(ts4) + length(ts4.pred)), ts4.pred, type = 'l', col = 2)  
  
#5. Write table and verify   
write.table(ts4.pred,  
 sep = ",",  
 col.names = FALSE,  
 row.names = FALSE,  
 file = "Q4\_Avery\_Kan\_25363807.txt")  
  
  
temp <- read.table("Q4\_Avery\_Kan\_25363807.txt", sep = ",")  
plot(as.numeric(unlist(temp)))  
  
  
####################################################################### DATASET 5

#1. Read in and explore  
ts5 <- read.csv("q5\_train.csv", as.is = TRUE)  
ts5 <- ts(ts5[,2])  
plot(ts5, type = 'l')  
#2. Infer from acf and pacf  
acf(ts5)  
ts5.d <- diff(ts5)  
ts5.acf <- ts5.d %>% acf(20)   
ts5.pacf <- ts5.d %>% pacf(20)  
ts5.acf[1]  
ts5.pacf[1]  
ts5.d.s <- ts5.d %>% diff(52)  
ts5.sacf <- ts5.d.s %>% acf(200)  
ts5.psacf <- ts5.d.s %>% pacf(200)  
ts5.sacf[52]   
ts5.psacf[52]  
#3. Model testing  
mse(c(3,1,2),c(0,1,1),ts5,length(ts5))   
#4.025  
mse(c(2,1,2),c(0,1,1),ts5,length(ts5))   
#4.0217  
mse(c(1,1,2),c(0,1,1),ts5,length(ts5))   
#4.0076  
  
##test MA  
mse(c(1,1,1),c(0,1,1),ts5,length(ts5))   
#3.9602  
#4. Predict and plot  
model <- arima(ts5, order = c(1,1,1), seasonal = list(order = c(0,1,1), period = 52))  
predict.arima <- predict(model,n.ahead= 104)  
ts5.pred <- predict.arima$pred  
plot(1:(length(ts5) + length(ts5.pred)), c(ts5, ts5.pred), type = 'l', col = 1)  
points((length(ts5) + 1) : (length(ts5) + length(ts5.pred)), ts5.pred, type = 'l', col = 2)  
  
#5. Write table and verify   
  
write.table(ts5.pred,  
 sep = ",",  
 col.names = FALSE,  
 row.names = FALSE,  
 file = "Q5\_Avery\_Kan\_25363807.txt")  
  
temp <- read.table("Q5\_Avery\_Kan\_25363807.txt", sep = ",")  
plot(as.numeric(unlist(temp)))