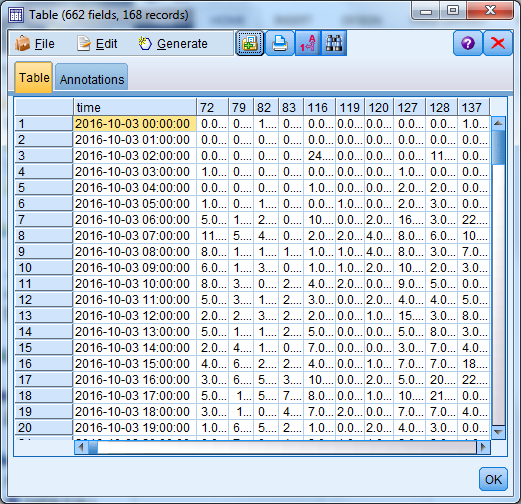
Exercise 10

The problem in this exercise uses data stored in the file *CitiBike\_Oct2016\_BikeDemand.csv*. The following view provides some details about the dataset.



Essentially the dataset provides you hourly based bike usage/demand information for selected 60 bike stations in New York City. Using moving average (k=30, 40, and 50) and exponential smoothing methods to answer the following questions:

* Question 1 (50 points): Choose the first 6 stations to predict their next 3 hours’ bike demands? Discuss the predicted results.

> movavg <- function(x, n, type=c("s", "t", "w", "m", "e", "r")) {

+ stopifnot(is.numeric(x), is.numeric(n), is.character(type))

+ if (length(n) != 1 || ceiling(n != floor(n)) || n <= 1)

+ stop("Window length 'n' must be a single integer greater 1.")

+ nx <- length(x)

+ if (n >= nx)

+ stop("Window length 'n' must be greater then length of time series.")

+ y <- numeric(nx)

+ if (type == "s") { # simple

+ for (k in 1:(n-1)) y[k] <- mean(x[1:k])

+ for (k in n:nx) y[k] <- mean(x[(k-n+1):k])

+ for (k in (nx+1):(nx+3)) y[k] <- mean((x+1)[(k-n+2):k])

+ } else if (type == "t") { # triangular

+ n <- ceiling((n + 1)/2)

+ s <- movavg(x, n, "s")

+ y <- movavg(s, n, "s")

+ } else if (type == "w") { # weighted

+ for (k in 1:(n-1)) y[k] <- 2 \* sum((k:1)\*x[k:1]) / (k\*(k+1))

+ for (k in n:nx) y[k] <- 2 \* sum((n:1)\*x[k:(k-n+1)]) / (n\*(n+1))

+ } else if (type == "m") { # modified

+ y[1] <- x[1]

+ for (k in 2:nx) y[k] <- y[k-1] + (x[k] - y[k-1])/n

+ } else if (type == "e") { # exponential

+ a <- 2/(n+1)

+ y[1] <- x[1]

+ for (k in 2:nx) y[k] <- a\*x[k] + (1-a)\*y[k-1]

+ for ( k in (nx+1): (nx+3)) y[k] <-a\*x[k] + (1-a)\*y[k-1]+a\*(x+1)[k]+(1-a)\*(a\*x[k] + (1-a)\*y[k-1])+a\*(x+2)[k]+(1-a)\*(a\*(x+1)[k]+(1-a)\*(a\*x[k] + (1-a)\*y[k-1]))

+ } else if (type == "r") { # running

+ a <- 1/n

+ y[1] <- x[1]

+ for (k in 2:nx) y[k] <- a\*x[k] + (1-a)\*y[k-1]

+ } else

+ stop("The type must be one of 's', 't', 'w', 'm', 'e', or 'r'.")

+ return(y)

+ }

> MyData <- read.csv(file="C:/Users/toddt/Downloads/CitiBike\_Oct2016\_BikeDemand.csv", header=TRUE, sep=",")

> bikestation<-MyData$X72

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation, 30, "s")

> lines(y, col = 2)

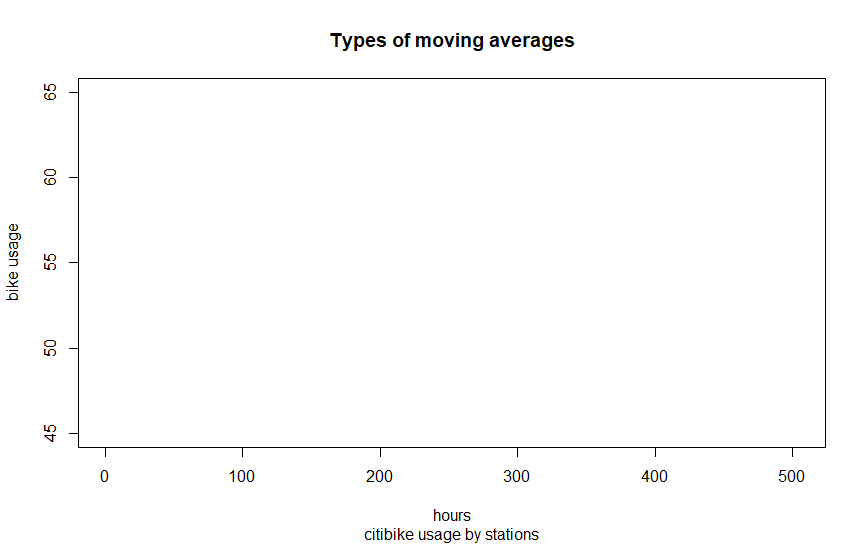
> y <- movavg(bikestation, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation, 50, "s")

> lines(y, col = 4)

>



> bikestation1<-MyData$X79

> plot(bikestation1, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation1, 30, "s")

> lines(y, col = 2)

> y <- movavg(bikestation1, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation1, 50, "s")

> y <- movavg(bikestation1, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(30)", "MV(40)", "MV(50)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation2<-MyData$X82

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation2, 30, "s")

> lines(y, col = 2)

> y <- movavg(bikestation2, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation2, 50, "s")

> y <- movavg(bikestation2, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(30)", "MV(40)", "MV(50)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation3<-MyData$X83

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation3, 30, "s")

> lines(y, col = 2)

> y <- movavg(bikestation3, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation3, 50, "s")

> y <- movavg(bikestation3, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(30)", "MV(40)", "MV(50)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation4<-MyData$X116

> plot(bikestation4, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation4, 30, "s")

> lines(y, col = 2)

> y <- movavg(bikestation4, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation4, 50, "s")

> y <- movavg(bikestation4, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(30)", "MV(40)", "MV(50)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation5<-MyData$X119

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation5, 30, "s")

> lines(y, col = 2)

> y <- movavg(bikestation5, 40, "s")

> lines(y, col = 3)

> y <- movavg(bikestation5, 50, "s")

> y <- movavg(bikestation5, 2, "e")

> lines(y, col = 6)

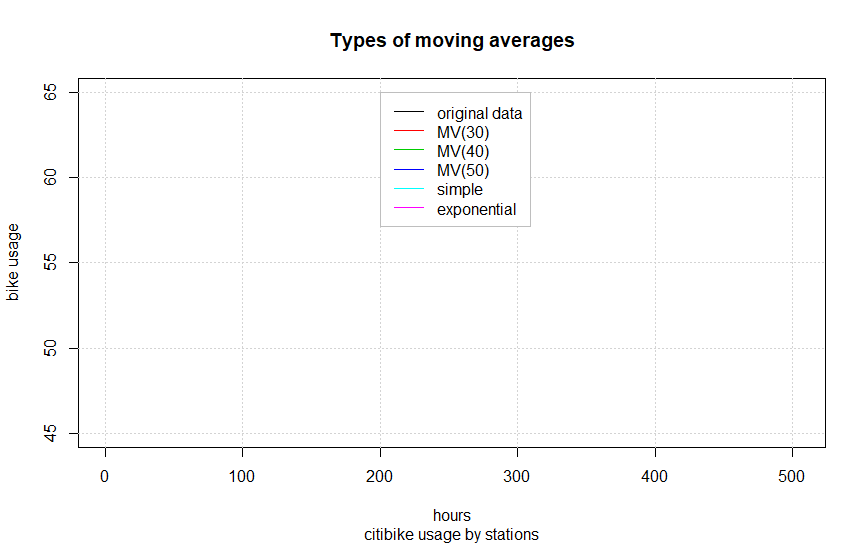
>

> grid()

>

> legend(200, 65, c("original data", "MV(30)", "MV(40)", "MV(50)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>



* Question 2 (50 points): Aggregating your hourly data into daily data, then choose the same 6 stations to predict their next three day’s bike demands (k adjusted to: k=5, 10, and 15)? Discuss the predicted results.

> library('data.table')

> DT<- data.table(MyData)

> DT[,sum(value),by = time]

> DT[,sum(X72),by = time]

time V1

1: 0010-03-20 75

2: 0010-04-20 90

3: 0010-05-20 106

4: 0010-06-20 93

5: 0010-07-20 88

6: 0010-08-20 64

7: 0010-09-20 36

8: 0010-10-20 91

9: 0010-11-20 88

10: 0010-12-20 67

11: <NA> 897

> dv<-DT[,sum(X72),by = time]

> View(dv)

> dv2<-DT[,sum(X79),by = time]

> dv3<-DT[,sum(X82),by = time]

> dv4<-DT[,sum(X83),by = time]

> dv5<-DT[,sum(X116),by = time]

> dv6<-DT[,sum(X119),by = time]

>

> bikestation<-DV$V1

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation, 15, "s")

> y <- movavg(bikestation, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation1<- DV2$V1

> plot(bikestation1, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation1, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation1, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation1, 15, "s")

> y <- movavg(bikestation1, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation2<- DV3$V1

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation2, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation2, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation2, 15, "s")

> y <- movavg(bikestation2, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation3<- DV4$V1

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation3, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation3, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation3, 15, "s")

> y <- movavg(bikestation3, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation4<- DV5$V1

> plot(bikestation4, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation4, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation4, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation4, 15, "s")

> y <- movavg(bikestation4, 2, "e")

> lines(y, col = 6)

>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>

> bikestation5<- DV6$V1

> plot(bikestation, type = "l", col = 1, ylim = c(45, 65), main = "Types of moving averages", sub = "citibike usage by stations", xlab = "hours", ylab = "bike usage")

> y <- movavg(bikestation5, 5, "s")

> lines(y, col = 2)

> y <- movavg(bikestation5, 10, "s")

> lines(y, col = 3)

> y <- movavg(bikestation5, 15, "s")

> y <- movavg(bikestation5, 2, "e")

> lines(y, col = 6)

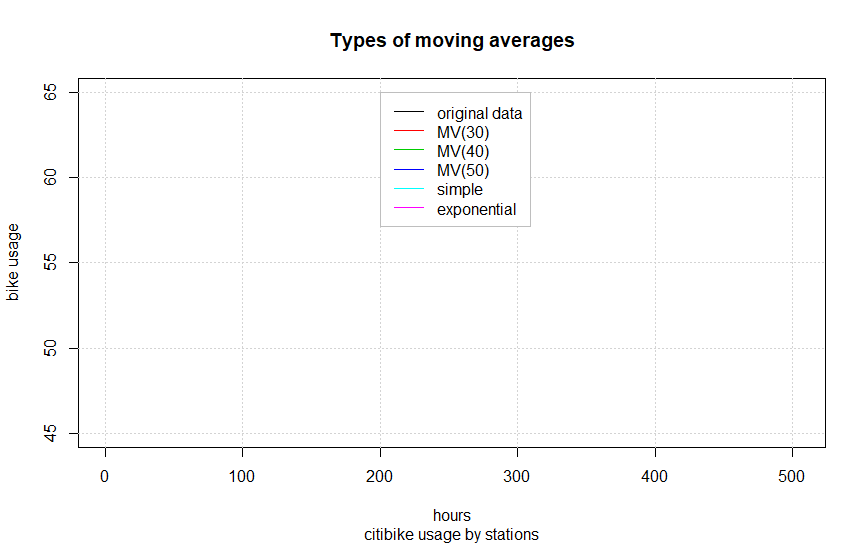
>

> grid()

>

> legend(200, 65, c("original data", "MV(5)", "MV(10)", "MV(15)", "simple", "exponential"), col = 1:7, lty = 1, lwd = 1, box.col = "gray", bg = "white")

>



**Tips:**

You need to modify the movavg function or define your own function based on the predicting functions listed below.

For moving average,

For exponential smoothing，

**Appendix** (The complete version of your solution scripts in R)