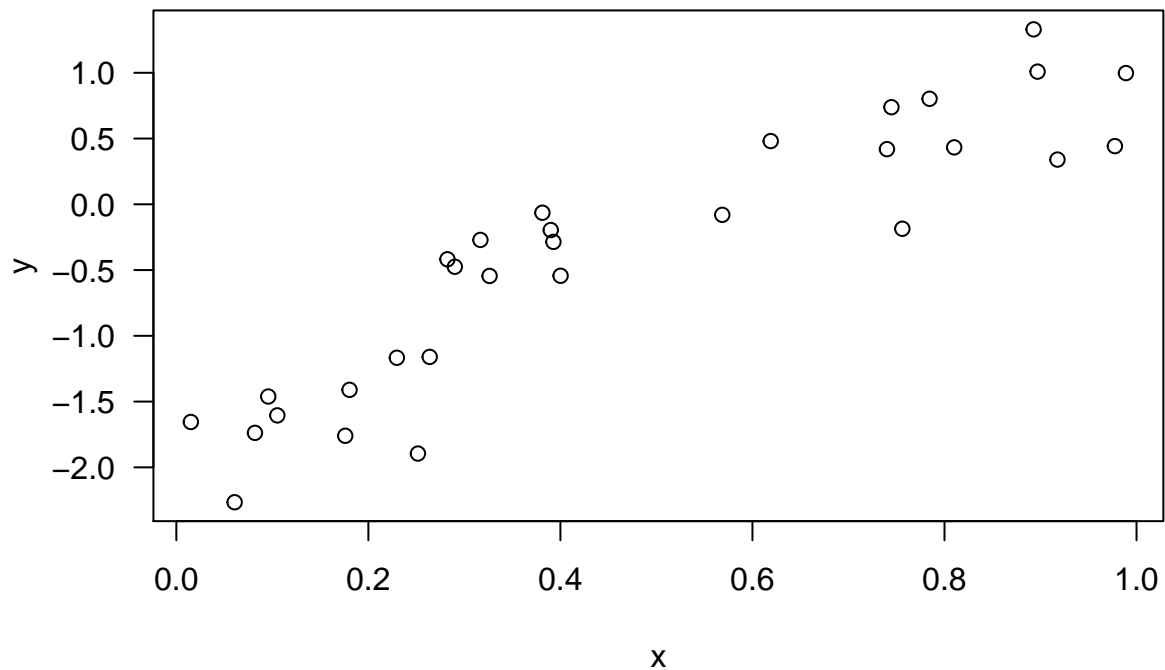


Creando-una-svm-manual-para-regresion.R

fhern

2020-09-23

```
# En este script se muestra como crear sum para regresion  
# de forma manual. Adicionalmente se comparan los resultados  
# con los obtenidos del funcion sum del paquete e1071  
  
# Generando unos datos artificiales  
gen_dat <- function (n) {  
  x <- runif(n)  
  y <- -2 + 3 * x + rnorm(n, sd=0.5)  
  datos <- data.frame(y=y, x=x)  
}  
  
#set.seed(2020)  
n <- 30  
datos <- gen_dat(n=n)  
  
# Explorando los datos  
with(datos, plot(x=x, y=y, las=1))
```



```
# Ajustando el modelo lm
mod1 <- lm(y ~ x, data=datos)
y_hat1 <- fitted(mod1)

# Mi sum manual -----

# Funcion objetivo a minimizar
func_obj_1 <- function(betas, C, my_epsilon, data) {
  b0 <- betas[1] # intercepto
  b1 <- betas[2] # pendiente
  y_hat <- b0 + b1 * data$x
  ei <- data$y - y_hat
  fuera <- abs(ei) > my_epsilon # identificando las obs fuera margen
  xi <- abs(ei[fuera]) - my_epsilon # los valores xi desde margen
  return(0.5*abs(b1)^2 + C * sum(xi))
}

# Usemos optim para encontrar los valores de b0 y b1
# que minimizan la funcion objetivo

my_epsilon <- 0.7
C <- 1

mod2 <- optim(par=c(0, 0), fn=func_obj_1,
              C=C, my_epsilon=my_epsilon, data=datos)
mod2
```

```
## $par
## [1] -1.395914  1.909881
##
## $value
## [1] 2.721844
##
## $counts
## function gradient
##      143      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
y_hat2 <- mod2$par[1] + mod2$par[2] * datos$x
```

```
# svm -----
library(e1071)
mod3 <- svm(y ~ x, data=datos,
            type="eps-regression",
            kernel="linear",
            scale=FALSE,
            cost=C,
            epsilon=my_epsilon)

y_hat3 <- predict(mod3, datos)

# Comparando las estimaciones
estimaciones <- cbind(lm=y_hat1,
                      svm_manual=y_hat2,
                      svm_e1071=y_hat3)
head(estimaciones, n=3)
```

```
##           lm  svm_manual  svm_e1071
## 1 -0.6249069 -0.65119096 -0.65119286
## 2  0.9219436  0.35668194  0.35668484
## 3  0.4014494  0.01754639  0.01754767
```

```
# Comparando los parametros
coef(mod1)
```

```
## (Intercept)          x
##   -1.767884    2.931223
```

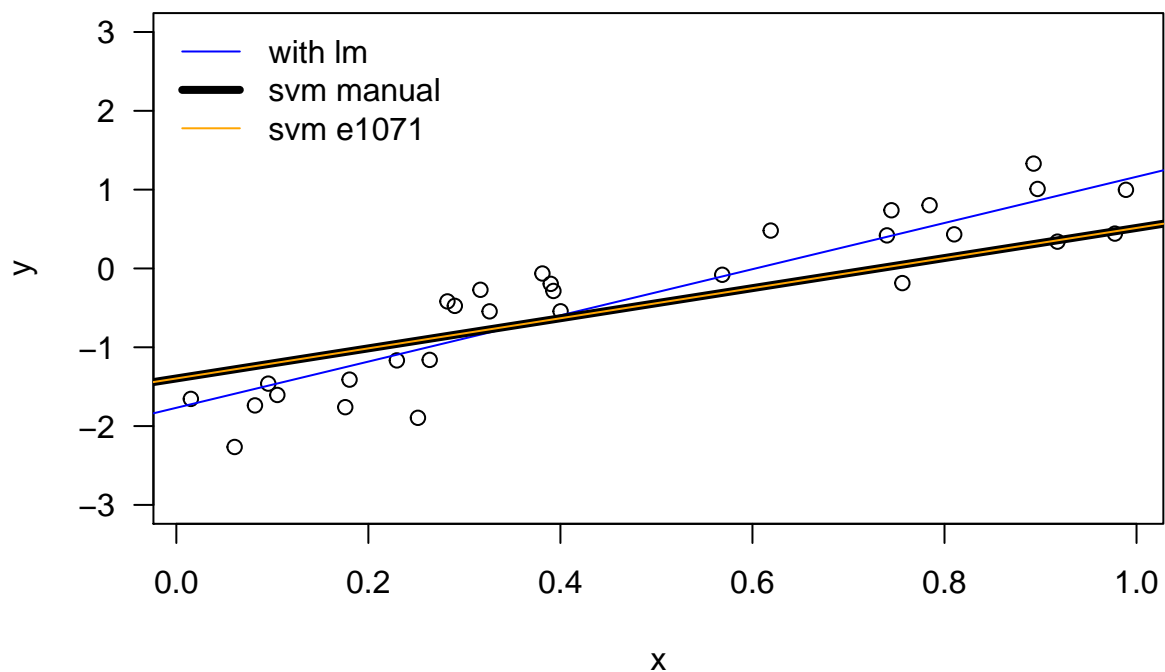
```
mod2$par
```

```
## [1] -1.395914  1.909881
```

```
coef(mod3)
```

```
## (Intercept)      x  
##    -1.39592    1.90989
```

```
# Agregando la recta al diagrama de dispersion  
with(datos, plot(x=x, y=y, las=1, ylim=c(-3, 3)))  
abline(mod1, col="blue")  
abline(a=mod2$par[1], b=mod2$par[2], col="black", lwd=4)  
abline(mod3, col="orange")  
  
legend("topleft", col=c("blue", "black", "orange"),  
      lwd=c(1, 4, 1), bty="n",  
      legend=c("with lm", "svm manual", "svm e1071"))
```



```
# Mostremos los margenes de svm -----  
with(datos, plot(x=x, y=y, las=1))  
abline(mod3)  
abline(a=coef(mod3)[1]-my_epsilon, b=coef(mod3)[2], lty="dashed")  
abline(a=coef(mod3)[1]+my_epsilon, b=coef(mod3)[2], lty="dashed")
```

