

```

rm(list = ls())
cat("\014")

set.seed(1)

train<-function(x,w,y){
  vtheta<-c()
  max_sum_xx<-c()
  m<-c()
  for (j in 1:ncol(x)) {
    xx<-cbind(x[,j],w,y)
    xx<-xx[order(xx[,1]),]
    sum_xx<-abs(cumsum(xx[,2]*xx[,3]))
    max_sum_xx[j]<-max(sum_xx)
    vtheta[j]<-xx[(which(sum_xx == max_sum_xx[j]))[1]],1]
  }
  J <- which(max_sum_xx==max(max_sum_xx))[1]
  m<-classify(x,list(j=J,theta=vtheta[J],m=m))
  if (length(which(y == m))<0.5*nrow(x)) m <- -m
  return(list(j=J,theta=vtheta[J],m=m))
}

classify<-function(x,pars) (sign(sign(x[,pars$j]-pars$theta)+0.1))

AdaBoost <- function(x,y,B){
  j <- c()
  theta <- c()
  alpha <- c()
  m<-matrix(nrow=nrow(x),ncol=B)
  w <- rep(1/nrow(x),nrow(x))
  for (b in 1:B) {
    pars<-train(x,w,y)
    theta[b]<-pars$theta
    j[b]<-pars$j
    m[,b]<-pars$m
    missc<-as.numeric((pars$m-y)!=0)
    alpha[b]<-log(-1+1/((w %*% missc)/sum(w)))
    w<-w*exp(alpha[b]*missc)
  }
  return(list(alpha=alpha,theta=theta,j=j,m=m))
}

agg_class <- function(x,alpha,allPars) {
  B <- length(allPars$j)
  m <- matrix(nrow = nrow(x), ncol = B)
  theta <- c()
  for (b in 1:B) {
    m[,b]<-classify(x,list(j=allPars$j[b],theta=allPars$theta[b],m=allPars$m[b]))
  }
  return(m)
}

misrate <- function(alpha,y,m) {

```

```

for (b in 1:length(alpha)) {
  if (length(which(y == m[,b])) < (0.5*length(y))) m[,b]<- -m[,b]
}
return(length(which(sign(m %*% alpha)!= y))/length(y))
}

x1<-read.csv("train.5.txt",header=FALSE)
x2<-read.csv("train.6.txt",header=FALSE)
x<-rbind(x1,x2)
y<-c(rep(1,nrow(x1)),rep(-1,nrow(x2)))

ii <- sample(1:nrow(x),nrow(x))
x<-x[ii,]
y<-y[ii]

B<-20
kfolds<-10

mistrain<-matrix(nrow=kfolds,ncol=B)
mistest<-matrix(nrow=kfolds,ncol=B)

fold<-sample(rep(1:kfolds,nrow(x)/kfolds))

for (contk in 1:kfolds) {

  ii<-which(fold %in% contk)

  ada <- AdaBoost(x[-ii,],y[-ii],B)

  for (b in 1:B) {
    allPars<-list(theta=ada$theta[1:b],j=ada$j[1:b],m=ada$m[,1:b])
    alpha<-ada$alpha[1:b]

    c_hat <- agg_class(x[-ii,],alpha,allPars)
    mistrain[contk,b]<-misrate(alpha,y[-ii],c_hat)

    c_hat <- agg_class(x[ii,],alpha,allPars)
    mistest[contk,b]<-misrate(alpha,y[ii],c_hat)
  }
}

vtrain<-apply(mistrain,2,mean)
vtest<-apply(mistest,2,mean)

plot(vtrain,type='o',col='blue',xlab = "Iterations", ylab = "Misclassification Rate")
grid(NULL)
lines(vtest, type='o',col = 'red')
legend("topright", legend = c('Testing error','Training error'), fill = c('red','blue'))

```

