set.seed(126)

n.points<-1000

age <- rpois(n.points, 50)

income<-rnorm(n.points,mean=50,sd=30)

credit <- sample(0:1, 1000, replace = TRUE)

is.numeric(credit)

is.factor(credit)

credit.f<-factor(credit, label=c("high","low"))

data=data.frame(age,income,credit.f)

head(data)

sampling.rate <-0.8

num.test.set.labels<- n.points\* (1 - sampling.rate)

training<-sample(1:n.points, sampling.rate\*n.points, replace=FALSE)

train<-subset(data[training,], select=c(age,income))

testing<-setdiff(1:n.points,training)

test<-subset(data[testing,], select=c(age,income))

cl<- data$credit.f[training]

true.labels<-data$credit.f[testing]

knn(train,test,cl,k=3) # we know class labels for train; Machine, Learn now.

#compute and print misclassification rate

for(k in 1:20) {

 print(k)

 predicted.labels <- knn(train,test,cl,k)

 num.incorrect.labels<-sum(predicted.labels !=true.labels)

 misclassification.rate<-num.incorrect.labels/num.test.set.labels

 print(misclassification.rate)

}

#now predict using k with lowest mis-classification rate

testcase<-c(57,37)

knn(train,testcase,cl,k=12)