

# kNN.R

*mac*

2019-11-08

```
Airb=read.csv("~/Documents/R/AB_NYC_2019.csv",header=T,na.strings="?")

attach(Airb)
library(sqldf)

## Loading required package: gsubfn
## Loading required package: proto
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 6): Library not loaded: /opt/X
##   Referenced from: /Library/Frameworks/R.framework/Versions/3.3/Resources/modules/R_X11.so
##   Reason: image not found
## Could not load tcltk. Will use slower R code instead.
## Loading required package: RSQLite
library(class)

price_x=Airb$price>100
Airb=data.frame(Airb,price_x)

TRain=sqldf("select longitude,latitude,minimum_nights,
            availability_365 from Airb where price!=0 and
            price<1000 and id%10!=1")

TEst=sqldf("select longitude,latitude,minimum_nights,
            availability_365 from Airb where price!=0 and
            price<1000 and id%10==1")

Train_label=(sqldf("select price_x from Airb where price!=0 and
                    price<1000 and id%10!=1"))

TEst_label=sqldf("select price_x from Airb where price!=0 and
                  price<1000 and id%10==1")

#normalization
TRain_z=as.data.frame(scale(TRain))
TEst_z=as.data.frame(scale(TEst))

pred=knn(train=TRain,test=TEst,cl=Train_label$price_x,k=20)

#result
table(Actual = TEst_label$price, Predicted = pred)

##          Predicted
## Actual  FALSE TRUE
```

```
## FALSE 1393 1028
## TRUE 725 1731

cm = as.matrix(table(Actual = TEst_label$price_x, Predicted = pred))
accu=sum(diag(cm))/length(TEst_label$price)
#accuracy
message(accu)

## 0.640557719909781
```