> source('~/.active-rstudio-document', echo=TRUE)

> library(class)

> library(ROCR)

> normalize <- function(x){

+ num <-x-min(x)

+ denom <- max(x)-min(x)

+ return (num/denom)

+ }

> Dset <- as.data.frame(lapply(train, normalize))

> summary(train)

ACTION RESOURCE MGR\_ID

Min. :0.0000 Min. : 0 Min. : 25

1st Qu.:1.0000 1st Qu.: 20299 1st Qu.: 4566

Median :1.0000 Median : 35376 Median : 13545

Mean :0.9421 Mean : 42924 Mean : 25989

3rd Qu.:1.0000 3rd Qu.: 74189 3rd Qu.: 42034

Max. :1.0000 Max. :312153 Max. :311696

ROLE\_ROLLUP\_1 ROLE\_ROLLUP\_2 ROLE\_DEPTNAME

Min. : 4292 Min. : 23779 Min. : 4674

1st Qu.:117961 1st Qu.:118102 1st Qu.:118395

Median :117961 Median :118300 Median :118921

Mean :116953 Mean :118302 Mean :118913

3rd Qu.:117961 3rd Qu.:118386 3rd Qu.:120535

Max. :311178 Max. :286791 Max. :286792

ROLE\_TITLE ROLE\_FAMILY\_DESC ROLE\_FAMILY

Min. :117879 Min. : 4673 Min. : 3130

1st Qu.:118274 1st Qu.:117906 1st Qu.:118363

Median :118568 Median :128696 Median :119006

Mean :125916 Mean :170178 Mean :183703

3rd Qu.:120006 3rd Qu.:235280 3rd Qu.:290919

Max. :311867 Max. :311867 Max. :308574

ROLE\_CODE

Min. :117880

1st Qu.:118232

Median :118570

Mean :119789

3rd Qu.:119348

Max. :270691

> responseY <- as.matrix(train[,dim(train)[2]])

> model.knn <- knn(train=train, test=train, cl=responseY, k=19, prob=T)

> plot(model.knn)

> predictedX <- predict(model, train)

> table(train$ACTION, predictedX > 0.5)

FALSE TRUE

0 4 1893

1 88 30784

> ROCRpred = prediction(predictedX, train$ACTION)

> auc = as.numeric(performance(ROCRpred, "auc")@y.values)

> ROCRperf = performance(ROCRpred, "tpr", "fpr")

> plot(ROCRperf)

> auc

[1] 0.4738415