Risk Classification

Mansoor Baba Shaik

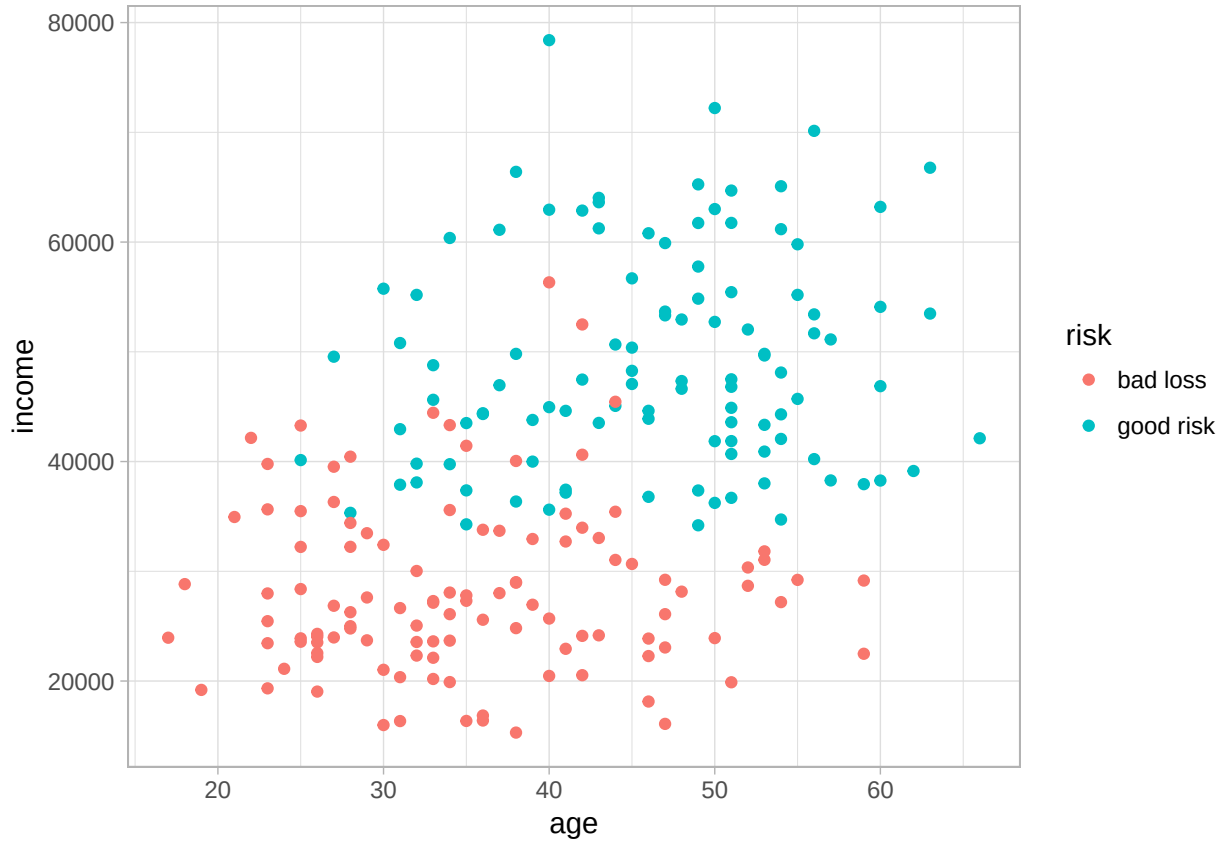
November 28, 2016

```
rm(list=ls())
setwd("~/IS665/risk")

classify_risk = read.csv("ClassifyRisk_historical.csv", header=TRUE, sep=",", dec=".")

require(ggplot2)

ggplot(data=classify_risk) + geom_point(aes(x=age, y=income, color=risk)) + theme_light()
```




```
require(class)
require(magrittr)

summary(classify_risk)
```

```
##  mortgage     loans            age         marital_status     income
##  n: 64    Min.   :0.000   Min.   :17.00   married:65     Min.   :15301
##  y:153    1st Qu.:1.000   1st Qu.:32.00   other  :53     1st Qu.:26857
##           Median :1.000   Median :40.00   single :99     Median :37363
##           Mean   :1.336   Mean   :40.54                  Mean   :38401
##           3rd Qu.:2.000   3rd Qu.:50.00                  3rd Qu.:47335
```

```
##           Max.   :3.000   Max.   :66.00                  Max.   :78399
##         risk
##  bad loss :111
##  good risk:106
##
##
##
##
```

```
n.classify_risk <- data.frame(sapply(classify_risk[,c(2,3,5)], function(x) {
(x - min(x))/(max(x) - min(x))
}))

head(n.classify_risk)
```

```
##       loans       age    income
## 1 0.3333333 0.3877551 0.2929044
## 2 0.3333333 0.7346939 0.4444643
## 3 0.3333333 0.9387755 0.6051203
## 4 0.3333333 0.6122449 0.6025608
## 5 0.6666667 0.6734694 0.3317385
## 6 0.0000000 0.6938776 0.7360926
```

```
# Split
set.seed(1234) # makes it repeatable
ind <- sample(2, nrow(classify_risk), replace = TRUE, prob = c(0.67, 0.33))
## Data Split
classify_risk.training <- n.classify_risk[ind == 1, 1:3]
classify_risk.test <- n.classify_risk[ind == 2, 1:3]

n.classify_risk[,"risk"] <- classify_risk[,"risk"]

## Label Split
classify_risk.trainLabels <- n.classify_risk[ind == 1, 4]
classify_risk.testLabels <- n.classify_risk[ind == 2, 4]
```

```
classify_risk_pred <- knn(train = classify_risk.training, test =
                          classify_risk.test, cl =
                          classify_risk.trainLabels,k = 5)

results = data.frame(classify_risk_pred, classify_risk.testLabels)
table(results)
```

```
##                   classify_risk.testLabels
## classify_risk_pred bad loss good risk
##          bad loss        27         1
##          good risk        2        32
```

```
# install.packages('caret')
require(caret)
# install.packages('e1071')
require(e1071)
confusionMatrix(table(results))
```

```
## Confusion Matrix and Statistics
##
##                   classify_risk.testLabels
```

```
## classify_risk_pred bad loss good risk
##          bad loss        27         1
##          good risk        2        32
##
##                Accuracy : 0.9516
##                  95% CI : (0.865, 0.9899)
##     No Information Rate : 0.5323
##     P-Value [Acc > NIR] : 2.844e-13
##
##                   Kappa : 0.9026
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9310
##             Specificity : 0.9697
##          Pos Pred Value : 0.9643
##          Neg Pred Value : 0.9412
##              Prevalence : 0.4677
##          Detection Rate : 0.4355
##    Detection Prevalence : 0.4516
##       Balanced Accuracy : 0.9504
##
##        'Positive' Class : bad loss
##
```

```
test_classify_risk <- read.csv("classifyrisk.csv", header=TRUE, sep=",", dec=".")

n.test_classify_risk <- data.frame(sapply(test_classify_risk[,c(2,3,5)], function(x) {
  (x - min(x))/(max(x) - min(x))
}))

classify_risk_pred.test <- n.test_classify_risk
head(classify_risk_pred.test)
```

```
##       loans       age    income
## 1 0.3333333 0.8139535 0.6121850
## 2 0.0000000 0.5813953 0.6242710
## 3 0.3333333 0.4418605 0.8957524
## 4 0.3333333 0.8372093 0.7006266
## 5 0.0000000 0.3720930 0.6875295
## 6 0.0000000 0.5581395 0.8846504
```

```
classify_test_result <- knn(train = classify_risk.training, test = classify_risk_pred.test,
                            cl = classify_risk[ind == 1, 6], k = 5)

(classify_test_result)
```

```
##  [1] good risk good risk good risk good risk good risk good risk bad loss
##  [8] bad loss  bad loss  good risk bad loss  good risk good risk good risk
## [15] good risk bad loss  good risk good risk bad loss  bad loss  good risk
## [22] good risk good risk good risk bad loss  bad loss  bad loss  good risk
## [29] bad loss
## Levels: bad loss good risk
```

```
test_classify_risk[,"risk"] <- classify_test_result

test_classify_risk
```

```
##    mortgage loans age marital_status   income      risk
## 1         n     1  54        married 50203.25 good risk
## 2         y     0  44        married 50793.46 good risk
## 3         y     1  38         single 64051.12 good risk
## 4         y     1  55        married 54522.25 good risk
## 5         y     0  35         single 53882.66 good risk
## 6         y     0  43         single 63508.96 good risk
## 7         n     2  62        married 27030.50  bad loss
## 8         y     3  41          other 20307.50  bad loss
## 9         y     1  26         single 24777.08  bad loss
## 10        y     0  50        married 48931.66 good risk
## 11        n     2  30          other 23575.22  bad loss
## 12        y     0  54        married 53242.06 good risk
## 13        n     1  48         single 39527.75 good risk
## 14        y     0  37        married 51852.40 good risk
## 15        y     1  43         single 69142.00 good risk
## 16        y     2  23          other 22039.40  bad loss
## 17        y     1  34        married 54180.75 good risk
## 18        y     1  54        married 53011.50 good risk
## 19        y     2  33         single 35558.00  bad loss
## 20        n     1  19        married 20954.50  bad loss
## 21        y     0  49        married 50873.56 good risk
## 22        y     2  48         single 39826.60 good risk
## 23        y     3  55          other 39417.88 good risk
## 24        y     1  38        married 52561.75 good risk
## 25        y     2  24         single 32626.00  bad loss
## 26        y     1  23         single 24314.50  bad loss
## 27        y     1  39         single 23785.64  bad loss
## 28        n     1  52        married 35380.75 good risk
## 29        n     1  49         single 29352.50  bad loss
```