Assignment 2 (Resubmission)

Jacob Aylward

2025-06-11

```
library(readxl)
UniversalBank <- read_excel("C:/Users/jacob/Downloads/UniversalBank.xlsx")
View(UniversalBank)

bank <- subset(UniversalBank, select = -c(ID, ZIPCode))

bank$Education <- factor(bank$Education)
library(fastDummies)
bank <- dummy_cols(bank, select_columns = "Education", remove_first_dummy =
FALSE, remove_selected_columns = TRUE)

set.seed(123)
n <- nrow(bank)
train_index <- sample(1:n, size = 0.6 * n)
train <- bank[train_index, ]
valid <- bank[-train_index, ]

norm_vars <- setdiff(names(bank), "PersonalLoan")
normalize <- function(x) { (x - min(x)) / (max(x) - min(x)) }
train_norm <- as.data.frame(lapply(train[, norm_vars], normalize))
valid_norm <- as.data.frame(lapply(valid[, norm_vars], normalize))
train_norm$PersonalLoan <- train$PersonalLoan
valid_norm$PersonalLoan <- valid$PersonalLoan

new_customer <- data.frame(
  Age = 40,
  Experience = 10,
  Income = 84,
  Family = 2,
  CCAvg = 2,
  Education_1 = 0,
  Education_2 = 1,
  Education_3 = 0,
  Mortgage = 0,
  SecuritiesAccount = 0,
  CDAccount = 0,
  Online = 1,
  CreditCard = 1)

for (col in names(new_customer)) {
  min_val <- min(train[[col]])
  max_val <- max(train[[col]])
  new_customer[[col]] <- (new_customer[[col]] - min_val) / (max_val -
```

```
min_val)
}

library(class)
k <- 1
pred <- knn(train = train_norm[, -which(names(train_norm) ==
"PersonalLoan")], test = new_customer, cl = train_norm$PersonalLoan, k = k)

print(pred)
```

```
## [1] 1
## Levels: 0 1
```