Assignment 3 - Universal Bank

Alexis McCartney

2025-10-13

Assignment 3 — Part A

```
# Knit-safe loader (no dialogs)
csv_path <- "UniversalBank.csv"
if (!file.exists(csv_path)) {
  stop("Missing file: UniversalBank.csv. Put the CSV in the same folder as this Rmd, then knit again.")
}
df <- read.csv(csv_path, check.names = TRUE, stringsAsFactors = FALSE)

# quick check
nrow(df); ncol(df)
```

```
## [1] 5000

## [1] 14
```

```
# Build pivot for Part A (rows: CC + Loan; cols: Online)
tab3 <- with(train, table(CC = CreditCard, Loan = Personal.Loan, Online = Online))
tab_df <- as.data.frame(tab3)  # CC, Loan, Online, Freq

pivot_wide <- reshape(tab_df,
                      idvar   = c("CC", "Loan"),
                      timevar = "Online",
                      direction = "wide")

# Tidy names and replace NAs with 0
names(pivot_wide)[names(pivot_wide) == "Freq.0"] <- "Online_0"
names(pivot_wide)[names(pivot_wide) == "Freq.1"] <- "Online_1"
if (!"Online_0" %in% names(pivot_wide)) pivot_wide$Online_0 <- 0L
if (!"Online_1" %in% names(pivot_wide)) pivot_wide$Online_1 <- 0L
pivot_wide$Online_0[is.na(pivot_wide$Online_0)] <- 0L
pivot_wide$Online_1[is.na(pivot_wide$Online_1)] <- 0L

# Order rows and print
pivot_wide <- pivot_wide[order(pivot_wide$CC, pivot_wide$Loan), ]
cat("\n== Pivot table (Training Set) ==\n")
```

```
##
## == Pivot table (Training Set) ==
```

```r
print(pivot_wide, row.names = FALSE)
```

```
##   CC Loan Online_0 Online_1
##    0    0      803     1125
##    0    1       77      131
##    1    0      312      472
##    1    1       32       48
```

```r
# Save pivot (optional)
out_path <- file.path(getwd(), "pivot_table_partA.csv")
write.csv(pivot_wide, out_path, row.names = FALSE)
cat("\nSaved pivot to: ", out_path, "\n", sep = "")
```

```
##
## Saved pivot to: /Users/alexismccartney/Desktop/BA64060/pivot_table_partA.csv
```

```r
# Interpretation
cat("\n**My Interpretation (Part A)**\n",
    "Online users (Online=1) appear more frequently across CC/Loan groups than offline users.\n",
    "Credit card holders who are online show slightly higher loan-acceptance counts.\n",
    sep = "")
```

```
##
## **My Interpretation (Part A)**
## Online users (Online=1) appear more frequently across CC/Loan groups than offline users.
## Credit card holders who are online show slightly higher loan-acceptance counts.
```

Assignment 3 — Part B

```r
cat("\n== Part B: Conditional Probability ==\n")
```

```
##
## == Part B: Conditional Probability ==
```

```r
# P(Loan=1 | CC=1, Online=1) from the pivot
cc1_rows <- subset(pivot_wide, CC == 1)
numerator <- cc1_rows$Online_1[cc1_rows$Loan == 1]
denominator <- sum(cc1_rows$Online_1)
prob_loan_given_cc1_online1 <- numerator / denominator

cat("P(Loan=1 | CC=1, Online=1) = ", round(prob_loan_given_cc1_online1, 4), "\n", sep = "")
```

```
## P(Loan=1 | CC=1, Online=1) = 0.0923
```

```r
cat("\n**My Interpretation (Part B)**\n",
    "Given CC=1 and Online=1, the sample conditional probability of accepting the loan is about ",
    round(100*prob_loan_given_cc1_online1, 2), "%.\n", sep = "")
```

```
##
## **My Interpretation (Part B)**
## Given CC=1 and Online=1, the sample conditional probability of accepting the loan is about 9.23%.
```

Assignment 3 — Part C

```r
cat("\n== Part C: Pivot Tables for Loan vs Online and Loan vs Credit Card ==\n")
```

```
##
## == Part C: Pivot Tables for Loan vs Online and Loan vs Credit Card ==
```

```r
pivot_loan_online <- with(train, table(Loan = Personal.Loan, Online = Online))
cat("\nPivot Table 1: Loan (rows) x Online (columns)\n")
```

```
##
## Pivot Table 1: Loan (rows) x Online (columns)
```

```r
print(pivot_loan_online)
```

```
##     Online
## Loan    0    1
##    0 1115 1597
##    1  109  179
```

```r
pivot_loan_cc <- with(train, table(Loan = Personal.Loan, CreditCard = CreditCard))
cat("\nPivot Table 2: Loan (rows) x CreditCard (columns)\n")
```

```
##
## Pivot Table 2: Loan (rows) x CreditCard (columns)
```

```r
print(pivot_loan_cc)
```

```
##     CreditCard
## Loan    0    1
##    0 1928  784
##    1  208   80
```

```r
write.csv(as.data.frame(pivot_loan_online),
          file.path(getwd(), "pivot_Loan_vs_Online.csv"),
          row.names = FALSE)
write.csv(as.data.frame(pivot_loan_cc),
          file.path(getwd(), "pivot_Loan_vs_CreditCard.csv"),
          row.names = FALSE)

cat("\n**My Interpretation (Part C)**\n",
    "Loan=1 counts are slightly higher among Online=1 and CC=1.\n", sep = "")
```

```
##
## **My Interpretation (Part C)**
## Loan=1 counts are slightly higher among Online=1 and CC=1.
```

Assignment 3 — Part D

```r
cat("\n== Part D: Conditional Probabilities ==\n")
```

```
##
## == Part D: Conditional Probabilities ==
```

```r
p_cc1_given_loan1    <- mean(train$CreditCard[train$Personal.Loan == 1] == 1)
p_online1_given_loan1<- mean(train$Online[train$Personal.Loan == 1] == 1)
p_loan1              <- mean(train$Personal.Loan == 1)

p_cc1_given_loan0    <- mean(train$CreditCard[train$Personal.Loan == 0] == 1)
p_online1_given_loan0<- mean(train$Online[train$Personal.Loan == 0] == 1)
p_loan0              <- mean(train$Personal.Loan == 0)

cat("\nP(CC=1 | Loan=1)  =", round(p_cc1_given_loan1, 4),
    "\nP(Online=1 | Loan=1) =", round(p_online1_given_loan1, 4),
    "\nP(Loan=1) =", round(p_loan1, 4),
    "\nP(CC=1 | Loan=0)  =", round(p_cc1_given_loan0, 4),
    "\nP(Online=1 | Loan=0) =", round(p_online1_given_loan0, 4),
    "\nP(Loan=0) =", round(p_loan0, 4), "\n")
```

```
##
## P(CC=1 | Loan=1)  = 0.2778
## P(Online=1 | Loan=1) = 0.6215
## P(Loan=1) = 0.096
## P(CC=1 | Loan=0)  = 0.2891
## P(Online=1 | Loan=0) = 0.5889
## P(Loan=0) = 0.904
```

```r
cat("\n**My Interpretation (Part D)**\n",
    "Acceptors have slightly higher Online=1; differences are modest overall.\n", sep = "")
```

```
##
## **My Interpretation (Part D)**
## Acceptors have slightly higher Online=1; differences are modest overall.
```

Assignment 3 — Part E (Naive Bayes)

```r
cat("\n== Part E: Naive Bayes P(Loan=1 | CC=1, Online=1) ==\n")
```

```
##
## == Part E: Naive Bayes P(Loan=1 | CC=1, Online=1) ==
```

```r
nb_num <- p_cc1_given_loan1 * p_online1_given_loan1 * p_loan1
nb_den <- nb_num + (p_cc1_given_loan0 * p_online1_given_loan0 * p_loan0)
nb_prob <- nb_num / nb_den

cat("Naive Bayes estimate = ", round(nb_prob, 4), "\n", sep = "")
```

```
## Naive Bayes estimate = 0.0972
```

```r
cat("\n**My Interpretation (Part E)**\n",
    "Under conditional independence, the NB probability is close to the sample conditional from Part B.\
```

```
##
## **My Interpretation (Part E)**
## Under conditional independence, the NB probability is close to the sample conditional from Part B.
```

Assignment 3 — Part F (Comparison)

```r
# Ensure the Part B value exists
if (!exists("prob_loan_given_cc1_online1")) {
  cc1_rows <- subset(pivot_wide, CC == 1)
  numerator <- cc1_rows$Online_1[cc1_rows$Loan == 1]
  denominator <- sum(cc1_rows$Online_1)
  prob_loan_given_cc1_online1 <- numerator / denominator
}

cat("\n== Part F: Compare Naive Bayes vs Empirical Probability ==\n")
```

```
##
## == Part F: Compare Naive Bayes vs Empirical Probability ==
```

```
cat("Empirical P(Loan=1 | CC=1, Online=1) =", round(prob_loan_given_cc1_online1, 4), "\n")
```

```
## Empirical P(Loan=1 | CC=1, Online=1) = 0.0923
```

```
cat("Naive Bayes P(Loan=1 | CC=1, Online=1) =", round(nb_prob, 4), "\n")
```

```
## Naive Bayes P(Loan=1 | CC=1, Online=1) = 0.0972
```

```
difference <- abs(nb_prob - prob_loan_given_cc1_online1)
cat("\nDifference =", round(difference, 4), "\n")
```

```
##
## Difference = 0.0049
```

```
cat("\n**My Interpretation (Part F)**\n",
    "The pivot-based estimate is the exact sample conditional; NB is a close model-based approximation.
    sep = "")
```

```
##
## **My Interpretation (Part F)**
## The pivot-based estimate is the exact sample conditional; NB is a close model-based approximation.
```

Assignment 3 — Part G (Naive Bayes Model)

```
cat("\n== Part G: Naive Bayes model & comparison ==\n")
```

```
##
## == Part G: Naive Bayes model & comparison ==
```

```
cat("\nEntries needed for P(Loan=1 | CC=1, Online=1):\n",
    "P(CC=1 | Loan=1), P(Online=1 | Loan=1), P(Loan=1),\n",
    "P(CC=1 | Loan=0), P(Online=1 | Loan=0), P(Loan=0)\n", sep="")
```

```
##
## Entries needed for P(Loan=1 | CC=1, Online=1):
## P(CC=1 | Loan=1), P(Online=1 | Loan=1), P(Loan=1),
## P(CC=1 | Loan=0), P(Online=1 | Loan=0), P(Loan=0)
```

```r
# Ensure e1071 is available
if (!requireNamespace("e1071", quietly = TRUE)) install.packages("e1071")
library(e1071)

# Fit NB model
nb_model <- naiveBayes(Personal.Loan ~ CreditCard + Online, data = train)
print(nb_model)
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##     0     1
## 0.904 0.096
##
## Conditional probabilities:
##    CreditCard
## Y           0         1
##   0 0.7109145 0.2890855
##   1 0.7222222 0.2777778
##
##    Online
## Y           0         1
##   0 0.4111357 0.5888643
##   1 0.3784722 0.6215278
```

```r
# Predict for CC=1, Online=1
new_customer <- data.frame(CreditCard = factor(1, levels = c(0,1)),
                           Online     = factor(1, levels = c(0,1)))
pred_nb <- predict(nb_model, new_customer, type = "raw")

cat("\nNB model P(Loan=1 | CC=1, Online=1) = ", round(pred_nb[,'1'], 4), "\n", sep = "")
```

```
##
## NB model P(Loan=1 | CC=1, Online=1) = 0.0972
```

```r
cat("\n**My Interpretation (Part G)**\n",
    "The model-based probability matches the manual NB calculation from Part E, confirming consistency.
    sep = "")
```

```
##
## **My Interpretation (Part G)**
## The model-based probability matches the manual NB calculation from Part E, confirming consistency.
```