

Tutorial 5 Memo

Dr. Niladri Chakraborty

2025-05-15

```
rm(list=ls())
# Install required packages if needed
library(AER)
library(e1071)

# 1. Load and preview the dataset
data("Affairs")
head(Affairs)

##   affairs gender age yearsmarried children religiousness education
##   occupation
## 4         0  male  37         10.00      no             3         18
## 7
## 5         0 female  27          4.00      no             4         14
## 6
## 11        0 female  32         15.00     yes             1         12
## 1
## 16         0  male  57         15.00     yes             5         18
## 6
## 23         0  male  22          0.75      no             2         17
## 6
## 29         0 female  32          1.50      no             2         17
## 5
##   rating
## 4         4
## 5         4
## 11        4
## 16         5
## 23         3
## 29         5

# 2. Create binary response variable
Affairs$had_affair <- as.factor(ifelse(Affairs$affairs > 0, 1, 0))
data = Affairs[, -1]
str(data)

## 'data.frame':   601 obs. of  9 variables:
##  $ gender      : Factor w/ 2 levels "female","male": 2 1 1 2 2 1 1 2 1 2
##  ...
##  $ age         : num  37 27 32 57 22 32 22 57 32 22 ...
##  $ yearsmarried : num  10 4 15 15 0.75 1.5 0.75 15 15 1.5 ...
##  $ children     : Factor w/ 2 levels "no","yes": 1 1 2 2 1 1 1 2 2 1 ...
##  $ religiousness: int   3 4 1 5 2 2 2 2 4 4 ...
##  $ education    : num  18 14 12 18 17 17 12 14 16 14 ...
```

```

## $ occupation   : int  7 6 1 6 6 5 1 4 1 4 ...
## $ rating       : int  4 4 4 5 3 5 3 4 2 5 ...
## $ had_affair   : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

# 3. Manual train-test split (70/30)
set.seed(100)
n <- nrow(data)
train_index <- sample(1:n, size = 0.7 * n)
train_data <- data[train_index, ]
test_data <- data[-train_index, ]

# 4. Train SVM model with linear kernel
svm_model <- svm(had_affair ~ .,
                 data = train_data, kernel = "linear", scale = TRUE,
                 probability = TRUE)

# 5. Predict on test data
pred_test <- predict(svm_model, newdata = test_data)

# 6. Calculate training error
training_error <- mean(pred_test != test_data$had_affair)
print(training_error)

## [1] 0.2099448

```

Conclusion: The model has approximately 21% error rate on prediction with unseen data indicating a moderately good model.