

# Tut6\_Memo

Dr. Niladri Chakraborty

2024-05-10

**Q1. Use the dataset 'Data\_Tutorial6.csv' from the 'Extra resource' section of the Blackboard. Scale the variables 'income', 'age', and 'balance' from the dataset.**

**(i) Create two clusters using the scaled data and create a cluster plot.**

**(ii) Use the full dataset to partition into training data and test data with 70:30 ratio. Then use the variables 'age' and 'income' as predictors and 'loan\_approved' as the target variable to fit a SVM model. Use a polynomial kernel.**

**(iii) Then create a SVM plot.**

Useful packages: tidyverse, lattice, caret, e1071, cluster

```
library(lattice)
library(caret)
library(tidyverse)
library(e1071)
library(cluster)
```

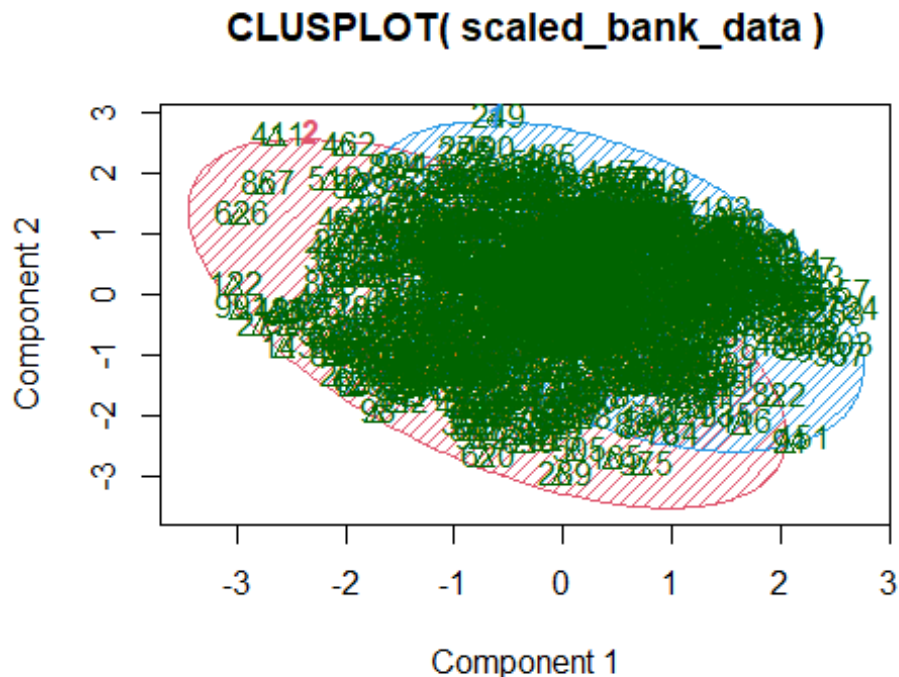
```
bank_data <- read.csv("C:\\Users\\ChakrabortyN\\OneDrive - University of the
Free
State\\Documents\\Admin_documents\\STSM2634_new_documents\\2024\\Datasets\\Da
ta_Tutorial6.csv", header =T)
```

```
scaled_bank_data <- as.data.frame(scale(bank_data[,c(1:3)]))
scaled_bank_data$loan_approved <- bank_data$loan_approved
```

```
bank_clusters <- kmeans(scaled_bank_data, centers = 2)
```

*# Plot the clusters*

```
clusplot(scaled_bank_data, bank_clusters$cluster, color=TRUE, shade=TRUE,
labels=2, lines=0)
```



These two components explain 53.22 % of the point variab

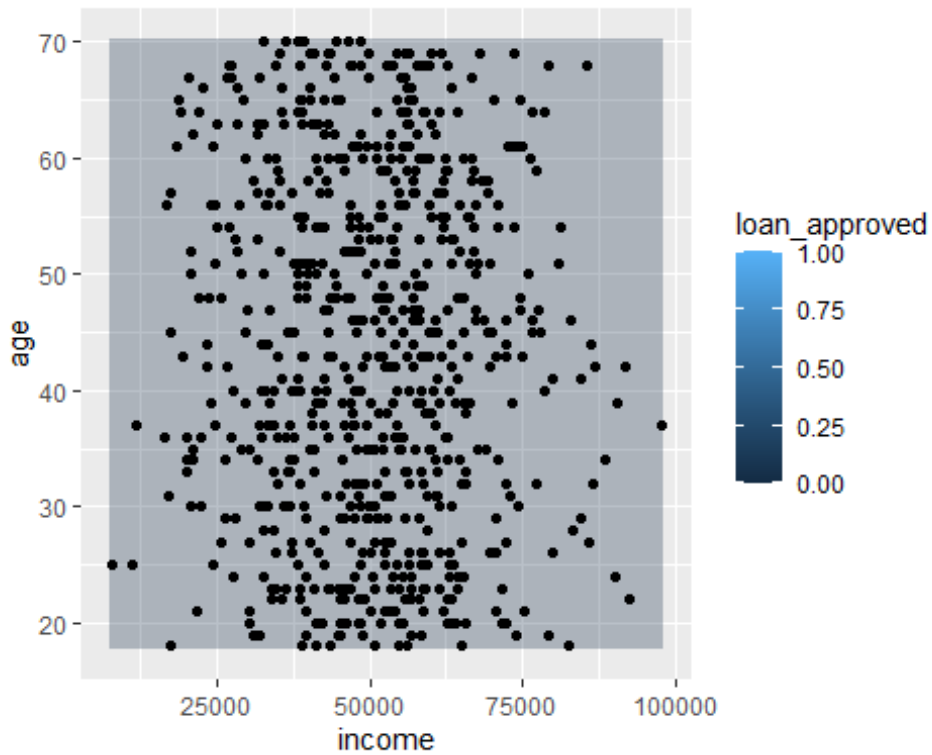
```
index <- createDataPartition(bank_data$loan_approved, p = 0.7, times = 1,
list = F)
data_train <- bank_data[index,]
data_test <- bank_data[-index,]
df <- data.frame(data_train[,c("age", "income", "loan_approved")])

svm_model <- svm(loan_approved ~ income+age, data = df, kernel =
'polynomial')

grid <- expand.grid(income = seq(min(df$income), max(df$income),
length.out = 100),
age = seq(min(df$age), max(df$age), length.out = 100))

grid$loan_approved <- predict(svm_model, newdata = grid)

ggplot(df, aes(income, age))+
  geom_tile(data = grid, aes(fill = loan_approved), alpha = 0.3)+
  geom_point(aes(fill = loan_approved))
```



```
df2 <- data.frame(data_test[,c("age", "income", "loan_approved")])
predict_loan <- predict(svm_model, df2)

postResample(predict_loan, df2$loan_approved)

##          RMSE      Rsquared        MAE
## 0.5268635996 0.0001110354 0.3213945195
```