# Read Dataset

library(readr)

zip\_test <- read\_delim("D:/UIC/Fall 2017/Advance stats/Assignments/Assignment1/zip.test.gz",

" ", escape\_double = FALSE, col\_names = FALSE,

trim\_ws = TRUE)

View(zip\_test)

library(readr)

zip\_train <- read\_delim("D:/UIC/Fall 2017/Advance stats/Assignments/Assignment1/zip.train.gz",

" ", escape\_double = FALSE, col\_names = FALSE,

trim\_ws = TRUE)

#View(zip\_train)

zip\_train <- zip\_train[,1:257] # cleaning the data

zipTrain <- subset(zip\_train,subset=X1%in%c(2,3)) #Considering 2's and 3's

#View(zipTrain)

zipTest <- subset(zip\_test,subset=X1%in%c(2,3)) #Considering 2's and 3's

#view(zipTest)

# Part 1: Linear Regresion

zip\_Model <- lm(X1 ~ ., zipTrain) # Linear Regression model

#print(zip\_Model$coefficients)

summary(zip\_Model)

par(mfrow = c(2,2))

plot(zip\_Model)

X1\_pred\_train<- predict(zip\_Model, zipTrain)

X1\_actual\_train = zipTrain$X1

X1\_train <- sqrt(mean((X1\_actual\_train - X1\_pred\_train)^2)) #RMSE Training

print(X1\_train)

X1\_predict<- predict(zip\_Model, zipTest)

X1\_actual = zipTest$X1

X1\_test <- sqrt(mean((X1\_actual - X1\_predict)^2)) #RMSE Test

print(X1\_test)

# Part 2: K-nearest neighbor

library(class)

knn\_1\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=1) # Knn model for K=1

knn\_1\_pred <- as.numeric(as.character(knn\_1\_train))

knn\_act = zipTrain$X1

table(knn\_1\_train,zipTrain$X1) # plotting confusion matrix for Knn=1 training data

RMSE\_1\_train <- mean(knn\_act != knn\_1\_pred) #Errro rate of Training for K=1

print(RMSE\_1\_train)

knn\_1<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=1) # Knn model for K=1

knn\_1\_predict <- as.numeric(as.character(knn\_1))

summary(knn\_1)

knn\_actual = zipTest$X1

table(knn\_1,zipTest$X1) # plotting confusion matrix for Knn=1 test data

RMSE\_1\_test <- mean(knn\_actual != knn\_1\_predict) #Errro rate of Test for K=1

print(RMSE\_1\_test)

knn\_3\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=3) # Knn model for K=3

str(knn\_3\_train)

knn\_3\_pred <- as.numeric(as.character(knn\_3\_train))

table(knn\_3\_train,zipTrain$X1) # plotting confusion matrix for Knn=3 training data

RMSE\_3\_train <- mean(knn\_act != knn\_3\_pred) #Errro rate of Training for K=3

print(RMSE\_3\_train)

knn\_3<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=3) # Knn model for K=3

knn\_3\_predict <- as.numeric(as.character(knn\_3))

table(knn\_3,zipTest$X1) # plotting confusion matrix for Knn=3 test data

RMSE\_3\_test <- mean(knn\_actual != knn\_3\_predict) #Errro rate of Test for K=3

print(RMSE\_3\_test)

knn\_5\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=5) # Knn model for K=5

knn\_5\_pred <- as.numeric(as.character(knn\_5\_train))

table(knn\_5\_train,zipTrain$X1) # plotting confusion matrix for Knn=5 training data

RMSE\_5\_train <- mean(knn\_act != knn\_5\_pred) #Errro rate of Training for K=5

print(RMSE\_5\_train)

knn\_5<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=5) # Knn model for K=5

knn\_5\_predict <- as.numeric(as.character(knn\_5))

table(knn\_5,zipTest$X1) # plotting confusion matrix for Knn=5 test data

RMSE\_5\_test <- mean(knn\_actual != knn\_5\_predict) #Errro rate of Test for K=5

print(RMSE\_5\_test)

knn\_7\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=7) # Knn model for K=7

knn\_7\_pred <- as.numeric(as.character(knn\_7\_train))

table(knn\_7\_train,zipTrain$X1) # plotting confusion matrix for Knn=7 training data

RMSE\_7\_train <- mean(knn\_act != knn\_7\_pred) #Errro rate of Training for K=7

print(RMSE\_7\_train)

knn\_7<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=7) # Knn model for K=7

knn\_7\_predict <- as.numeric(as.character(knn\_7))

table(knn\_7,zipTest$X1) # plotting confusion matrix for Knn=7 test data

RMSE\_7\_test <- mean(knn\_actual != knn\_7\_predict) #Errro rate of Test for K=7

print(RMSE\_7\_test)

knn\_15\_train<-knn(train=zipTrain[,2:257],test=zipTrain[,2:257],cl = as.factor(zipTrain$X1),k=15) # Knn model for K=15

knn\_15\_pred <- as.numeric(as.character(knn\_15\_train))

table(knn\_15\_train,zipTrain$X1) # plotting confusion matrix for Knn=15 training data

RMSE\_15\_train <- mean(knn\_act != knn\_15\_pred) #Errro rate of Training for K=15

print(RMSE\_15\_train)

knn\_15<-knn(train=zipTrain[,2:257],test=zipTest[,2:257],cl = as.factor(zipTrain$X1),k=15) # Knn model for K=15

knn\_15\_predict <- as.numeric(as.character(knn\_15))

table(knn\_15,zipTest$X1) # plotting confusion matrix for Knn=15 test data

RMSE\_15\_test <- mean(knn\_actual != knn\_15\_predict) #Errro rate of Test for K=7

print(RMSE\_15\_test)