Assignment 8.1

Problem Statement

1. Use the below given data set

DataSet

a. Predict the no of comments in next H hrs

b. Use regression technique

c. Report the training accuracy and test accuracy

Ans

library(C50)

data(churn)

head(churnTrain)

> head(churnTrain)

state account\_length area\_code international\_plan voice\_mail\_plan number\_vmail\_messages

1 KS 128 area\_code\_415 no yes 25

2 OH 107 area\_code\_415 no yes 26

3 NJ 137 area\_code\_415 no no 0

4 OH 84 area\_code\_408 yes no 0

5 OK 75 area\_code\_415 yes no 0

6 AL 118 area\_code\_510 yes no 0

total\_day\_minutes total\_day\_calls total\_day\_charge total\_eve\_minutes total\_eve\_calls

1 265.1 110 45.07 197.4 99

2 161.6 123 27.47 195.5 103

3 243.4 114 41.38 121.2 110

4 299.4 71 50.90 61.9 88

5 166.7 113 28.34 148.3 122

6 223.4 98 37.98 220.6 101

total\_eve\_charge total\_night\_minutes total\_night\_calls total\_night\_charge total\_intl\_minutes

1 16.78 244.7 91 11.01 10.0

2 16.62 254.4 103 11.45 13.7

3 10.30 162.6 104 7.32 12.2

4 5.26 196.9 89 8.86 6.6

5 12.61 186.9 121 8.41 10.1

6 18.75 203.9 118 9.18 6.3

head(churnTest)

> head(churnTest)

state account\_length area\_code international\_plan voice\_mail\_plan number\_vmail\_messages

1 HI 101 area\_code\_510 no no 0

2 MT 137 area\_code\_510 no no 0

3 OH 103 area\_code\_408 no yes 29

4 NM 99 area\_code\_415 no no 0

5 SC 108 area\_code\_415 no no 0

6 IA 117 area\_code\_415 no no 0

total\_day\_minutes total\_day\_calls total\_day\_charge total\_eve\_minutes total\_eve\_calls

1 70.9 123 12.05 211.9 73

2 223.6 86 38.01 244.8 139

3 294.7 95 50.10 237.3 105

4 216.8 123 36.86 126.4 88

5 197.4 78 33.56 124.0 101

6 226.5 85 38.51 141.6 68

total\_eve\_charge total\_night\_minutes total\_night\_calls total\_night\_charge total\_intl\_minutes

1 18.01 236.0 73 10.62 10.6

2 20.81 94.2 81 4.24 9.5

3 20.17 300.3 127 13.51 13.7

4 10.74 220.6 82 9.93 15.7

5 10.54 204.5 107 9.20 7.7

6 12.04 223.0 90 10.04 6.9

#churnTrain = churnTrain[1:500,]

#churnTest = churnTest[1:500,]

# logistic regression model:

fit <- glm(churn~.,data = churnTrain,family = binomial(link='logit'))

> fit <- glm(churn~.,data = churnTrain,family = binomial(link='logit'))

> summary(fit)

Call:

glm(formula = churn ~ ., family = binomial(link = "logit"), data = churnTrain)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.0431 0.1661 0.3123 0.4995 1.9487

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 9.686e+00 9.798e-01 9.885 < 2e-16 \*\*\*

stateAL -3.385e-01 7.629e-01 -0.444 0.657272

stateAR -9.106e-01 7.519e-01 -1.211 0.225884

stateAZ -8.973e-02 8.452e-01 -0.106 0.915453

stateCA -1.816e+00 7.822e-01 -2.322 0.020238 \*

stateCO -6.445e-01 7.631e-01 -0.845 0.398339

stateCT -1.021e+00 7.252e-01 -1.408 0.159167

stateDC -6.880e-01 8.081e-01 -0.851 0.394577

stateDE -7.460e-01 7.490e-01 -0.996 0.319234

stateFL -5.916e-01 7.610e-01 -0.777 0.436956

stateGA -6.601e-01 7.778e-01 -0.849 0.396075

stateHI 2.300e-01 8.963e-01 0.257 0.797469

stateIA -2.083e-01 9.024e-01 -0.231 0.817410

stateID -8.705e-01 7.474e-01 -1.165 0.244100

stateIL 2.382e-01 8.340e-01 0.286 0.775165

stateIN -4.410e-01 7.526e-01 -0.586 0.557924

stateKS -1.062e+00 7.296e-01 -1.455 0.145659

stateKY -7.889e-01 7.658e-01 -1.030 0.302931

stateLA -5.546e-01 8.352e-01 -0.664 0.506716

stateMA -1.161e+00 7.430e-01 -1.562 0.118261

stateMD -1.144e+00 7.168e-01 -1.596 0.110430

stateME -1.327e+00 7.281e-01 -1.823 0.068321 .

stateMI -1.390e+00 7.137e-01 -1.948 0.051400 .

stateMN -1.160e+00 7.149e-01 -1.622 0.104709

stateMO -5.979e-01 7.741e-01 -0.772 0.439914

stateMS -1.355e+00 7.278e-01 -1.862 0.062601 .

stateMT -1.865e+00 7.166e-01 -2.603 0.009245 \*\*

stateNC -5.765e-01 7.545e-01 -0.764 0.444822

stateND -1.274e-01 7.969e-01 -0.160 0.872995

stateNE -2.952e-01 8.055e-01 -0.367 0.713984

stateNH -1.160e+00 7.689e-01 -1.509 0.131367

stateNJ -1.572e+00 7.098e-01 -2.215 0.026757 \*

stateNM -4.590e-01 7.867e-01 -0.583 0.559596

stateNV -1.251e+00 7.245e-01 -1.727 0.084198 .

stateNY -1.161e+00 7.191e-01 -1.614 0.106496

stateOH -6.726e-01 7.464e-01 -0.901 0.367508

stateOK -8.660e-01 7.557e-01 -1.146 0.251811

stateOR -7.684e-01 7.354e-01 -1.045 0.296126

statePA -1.141e+00 7.791e-01 -1.464 0.143121

stateRI 1.099e-01 8.198e-01 0.134 0.893337

stateSC -1.747e+00 7.371e-01 -2.370 0.017782 \*

stateSD -8.227e-01 7.607e-01 -1.081 0.279510

stateTN -2.604e-01 8.207e-01 -0.317 0.751071

stateTX -1.637e+00 7.079e-01 -2.313 0.020745 \*

stateUT -1.047e+00 7.435e-01 -1.408 0.159056

stateVA 4.425e-01 8.220e-01 0.538 0.590344

stateVT -8.390e-02 7.799e-01 -0.108 0.914330

stateWA -1.400e+00 7.237e-01 -1.934 0.053081 .

stateWI -2.836e-01 7.798e-01 -0.364 0.716109

stateWV -5.732e-01 7.329e-01 -0.782 0.434139

stateWY -2.952e-01 7.541e-01 -0.391 0.695449

account\_length -9.646e-04 1.434e-03 -0.673 0.501212

area\_codearea\_code\_415 7.876e-02 1.418e-01 0.555 0.578569

area\_codearea\_code\_510 1.016e-01 1.632e-01 0.622 0.533622

international\_planyes -2.192e+00 1.534e-01 -14.294 < 2e-16 \*\*\*

voice\_mail\_planyes 2.131e+00 5.944e-01 3.585 0.000337 \*\*\*

number\_vmail\_messages -3.832e-02 1.865e-02 -2.055 0.039866 \*

total\_day\_minutes 3.823e-01 3.380e+00 0.113 0.909942

total\_day\_calls -4.045e-03 2.862e-03 -1.414 0.157477

total\_day\_charge -2.326e+00 1.988e+01 -0.117 0.906870

total\_eve\_minutes -8.927e-01 1.700e+00 -0.525 0.599510

total\_eve\_calls -1.018e-03 2.890e-03 -0.352 0.724642

total\_eve\_charge 1.041e+01 2.000e+01 0.521 0.602695

total\_night\_minutes 2.228e-01 9.044e-01 0.246 0.805401

total\_night\_calls -1.810e-04 2.928e-03 -0.062 0.950718

total\_night\_charge -5.039e+00 2.010e+01 -0.251 0.802042

total\_intl\_minutes 4.149e+00 5.494e+00 0.755 0.450194

total\_intl\_calls 9.055e-02 2.575e-02 3.516 0.000438 \*\*\*

total\_intl\_charge -1.567e+01 2.035e+01 -0.770 0.441115

number\_customer\_service\_calls -5.366e-01 4.100e-02 -13.089 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2758.3 on 3332 degrees of freedom

Residual deviance: 2070.8 on 3263 degrees of freedom

AIC: 2210.8

Number of Fisher Scoring iterations: 6



