17

Problem Statement

1. Use the below given data set

2. Perform the below given activities:

a. Create classification model using logistic regression model

> predict<- predict(model, type = "response")

> head(predict, 5)

1 2 3 4 5

0.9604327 0.9608404 0.9589231 0.9600989 0.9607629

> input$predict<- predict

> input$predictROUND<- round(predict, digits = 0)

> table(input$new\_window, predict>= 0.5)

TRUE

no 3936

yes 88

> dim(input)

[1] 4024 161

b. verify model goodness of fit

c. Report the accuracy measures

f. interpret the results

g. visualize the results

for questions (b,c,f,g) – Ans is as below

> model<- glm(input1~raw\_timestamp\_part\_1+raw\_timestamp\_part\_2+cvtd\_timestamp+num\_window+roll\_belt+pitch\_belt+yaw\_belt+total\_accel\_belt,data = input)

> model

Call: glm(formula = input1 ~ raw\_timestamp\_part\_1 + raw\_timestamp\_part\_2 +

cvtd\_timestamp + num\_window + roll\_belt + pitch\_belt + yaw\_belt +

total\_accel\_belt, data = input)

Coefficients:

(Intercept) raw\_timestamp\_part\_1 raw\_timestamp\_part\_2 cvtd\_timestamp05-12-2011 11:23 cvtd\_timestamp05-12-2011 11:25

-9.841e+05 7.440e-04 1.242e-07 -1.869e+02 -1.870e+02

cvtd\_timestamp05-12-2011 14:22 cvtd\_timestamp05-12-2011 14:23 cvtd\_timestamp28-11-2011 14:15 cvtd\_timestamp30-11-2011 17:12 num\_window

-1.949e+02 -1.949e+02 2.554e+02 1.192e+02 -8.201e-04

roll\_belt pitch\_belt yaw\_belt total\_accel\_belt

-4.217e-04 -4.897e-04 9.792e-05 2.525e-03

Degrees of Freedom: 4023 Total (i.e. Null); 4010 Residual

Null Deviance: 86.08

Residual Deviance: 80.79 AIC: -4277

> summary(model)

Call:

glm(formula = input1 ~ raw\_timestamp\_part\_1 + raw\_timestamp\_part\_2 +

cvtd\_timestamp + num\_window + roll\_belt + pitch\_belt + yaw\_belt +

total\_accel\_belt, data = input)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.25039 -0.04901 -0.01883 0.01123 0.96934

Coefficients:

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

(Intercept) -9.841e+05 4.774e+06 -0.206 0.837

raw\_timestamp\_part\_1 7.440e-04 3.609e-03 0.206 0.837

raw\_timestamp\_part\_2 1.242e-07 7.888e-09 15.747 <2e-16 \*\*\*

cvtd\_timestamp05-12-2011 11:23 -1.869e+02 9.069e+02 -0.206 0.837

cvtd\_timestamp05-12-2011 11:25 -1.870e+02 9.072e+02 -0.206 0.837

cvtd\_timestamp05-12-2011 14:22 -1.949e+02 9.455e+02 -0.206 0.837

cvtd\_timestamp05-12-2011 14:23 -1.949e+02 9.455e+02 -0.206 0.837

cvtd\_timestamp28-11-2011 14:15 2.554e+02 1.239e+03 0.206 0.837

cvtd\_timestamp30-11-2011 17:12 1.192e+02 5.766e+02 0.207 0.836

num\_window -8.201e-04 4.223e-03 -0.194 0.846

roll\_belt -4.217e-04 5.029e-04 -0.839 0.402

pitch\_belt -4.897e-04 1.151e-03 -0.426 0.670

yaw\_belt 9.792e-05 1.168e-04 0.839 0.402

total\_accel\_belt 2.525e-03 1.896e-03 1.332 0.183

---

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

(Dispersion parameter for gaussian family taken to be 0.02014796)

Null deviance: 86.076 on 4023 degrees of freedom

Residual deviance: 80.793 on 4010 degrees of freedom

AIC: -4276.7

Number of Fisher Scoring iterations: 2

d. Report the variable importance

Ans: variables highlighted in pink are important variables

e. Report the unimportant variables

Ans: variables highlighted in green are unimportant variables