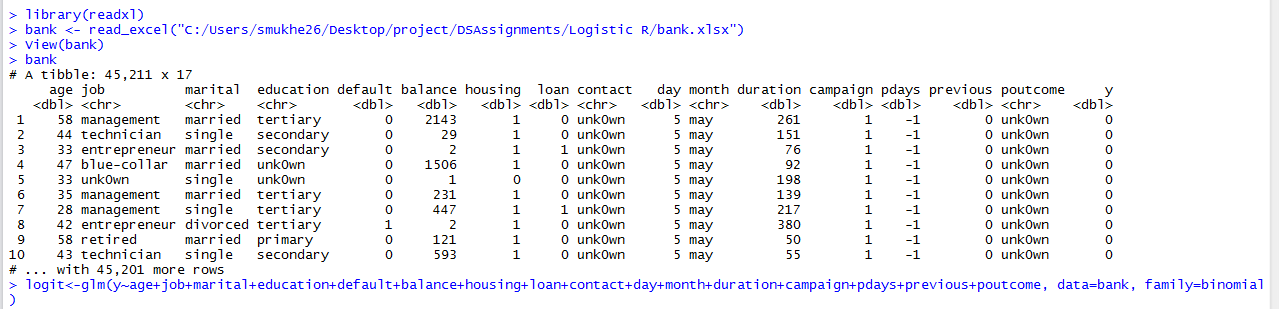
Output variable -> y

y -> Whether the client has subscribed a term deposit or not

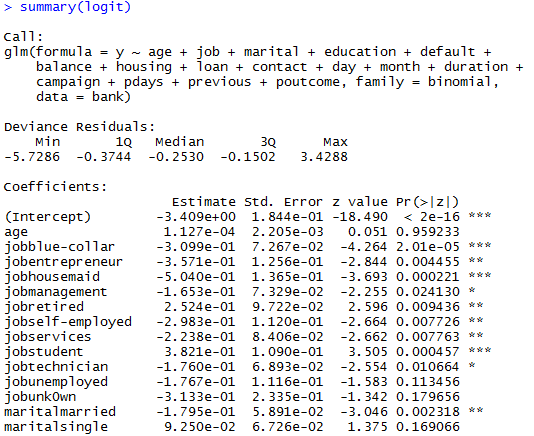
Binomial ("yes" or "no")

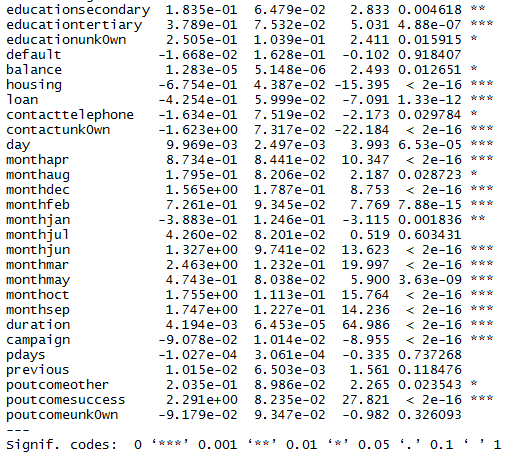
The data is in csv format delimited wit ;

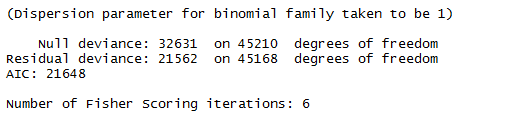
1. Import the data into xls format
2. Convert the variable Y from categorical to numerical (Yes=0 and no to 1) since logistic regression requires the target data to be in numerical
3. Import the dataset in R studio
4. Apply logistic regression on all the variable except the target



Do a suumary of logit to find the details







We can see that the difference in NULL Deviance and Residual Deviance is good and we can assume for now that the model is good.

But in order to perform better we should select the optimum variables which are helpful to predict the target output

Hence use stepAIC

> library(MASS)

> stepAIC(logit)

Start: AIC=21648.27

y ~ age + job + marital + education + default + balance + housing +

loan + contact + day + month + duration + campaign + pdays +

previous + poutcome

Df Deviance AIC

- age 1 21562 21646

- default 1 21562 21646

- pdays 1 21562 21646

<none> 21562 21648

- previous 1 21565 21649

- balance 1 21568 21652

- day 1 21578 21662

- education 3 21590 21670

- marital 2 21601 21683

- loan 1 21616 21700

- job 11 21659 21723

- campaign 1 21657 21741

- housing 1 21804 21888

- contact 2 22102 22184

- month 11 22600 22664

- poutcome 3 22729 22809

- duration 1 27152 27236

Step: AIC=21646.27

y ~ job + marital + education + default + balance + housing +

loan + contact + day + month + duration + campaign + pdays +

previous + poutcome

Df Deviance AIC

- default 1 21562 21644

- pdays 1 21562 21644

<none> 21562 21646

- previous 1 21565 21647

- balance 1 21568 21650

- day 1 21578 21660

- education 3 21590 21668

- marital 2 21605 21685

- loan 1 21616 21698

- job 11 21663 21725

- campaign 1 21657 21739

- housing 1 21807 21889

- contact 2 22103 22183

- month 11 22600 22662

- poutcome 3 22730 22808

- duration 1 27153 27235

Step: AIC=21644.28

y ~ job + marital + education + balance + housing + loan + contact +

day + month + duration + campaign + pdays + previous + poutcome

Df Deviance AIC

- pdays 1 21562 21642

<none> 21562 21644

- previous 1 21565 21645

- balance 1 21568 21648

- day 1 21578 21658

- education 3 21590 21666

- marital 2 21605 21683

- loan 1 21616 21696

- job 11 21663 21723

- campaign 1 21657 21737

- housing 1 21807 21887

- contact 2 22103 22181

- month 11 22601 22661

- poutcome 3 22730 22806

- duration 1 27153 27233

Step: AIC=21642.4

y ~ job + marital + education + balance + housing + loan + contact +

day + month + duration + campaign + previous + poutcome

Df Deviance AIC

<none> 21562 21642

- previous 1 21565 21643

- balance 1 21569 21647

- day 1 21578 21656

- education 3 21590 21664

- marital 2 21606 21682

- loan 1 21616 21694

- job 11 21663 21721

- campaign 1 21657 21735

- housing 1 21809 21887

- contact 2 22103 22179

- month 11 22602 22660

- poutcome 3 22783 22857

- duration 1 27153 27231

Call: glm(formula = y ~ job + marital + education + balance + housing +

loan + contact + day + month + duration + campaign + previous +

poutcome, family = binomial, data = bank)

Coefficients:

(Intercept) jobblue-collar jobentrepreneur jobhousemaid jobmanagement jobretired jobself-employed

-3.426e+00 -3.103e-01 -3.573e-01 -5.028e-01 -1.652e-01 2.552e-01 -2.981e-01

jobservices jobstudent jobtechnician jobunemployed jobunk0wn maritalmarried maritalsingle

-2.241e-01 3.819e-01 -1.758e-01 -1.771e-01 -3.124e-01 -1.792e-01 9.171e-02

educationsecondary educationtertiary educationunk0wn balance housing loan contacttelephone

1.832e-01 3.790e-01 2.506e-01 1.289e-05 -6.767e-01 -4.259e-01 -1.629e-01

contactunk0wn day monthapr monthaug monthdec monthfeb monthjan

-1.622e+00 9.976e-03 8.706e-01 1.775e-01 1.563e+00 7.248e-01 -3.899e-01

monthjul monthjun monthmar monthmay monthoct monthsep duration

4.005e-02 1.325e+00 2.462e+00 4.705e-01 1.753e+00 1.745e+00 4.194e-03

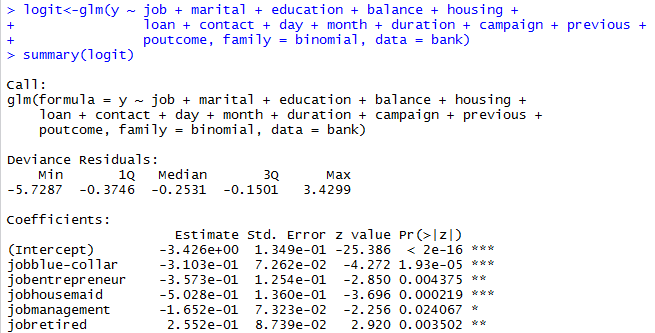
campaign previous poutcomeother poutcomesuccess poutcomeunk0wn

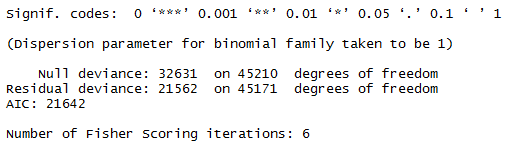
-9.082e-02 1.022e-02 2.049e-01 2.298e+00 -6.803e-02

Degrees of Freedom: 45210 Total (i.e. Null); 45171 Residual

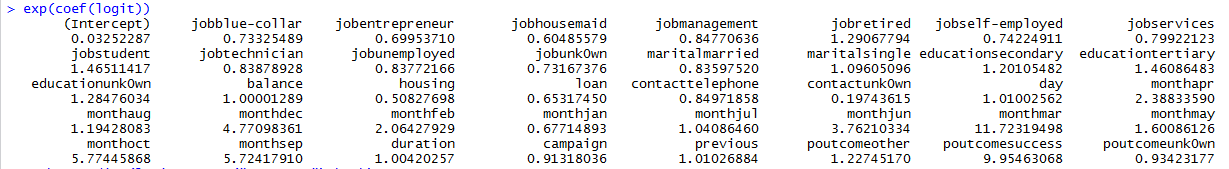
Null Deviance: 32630

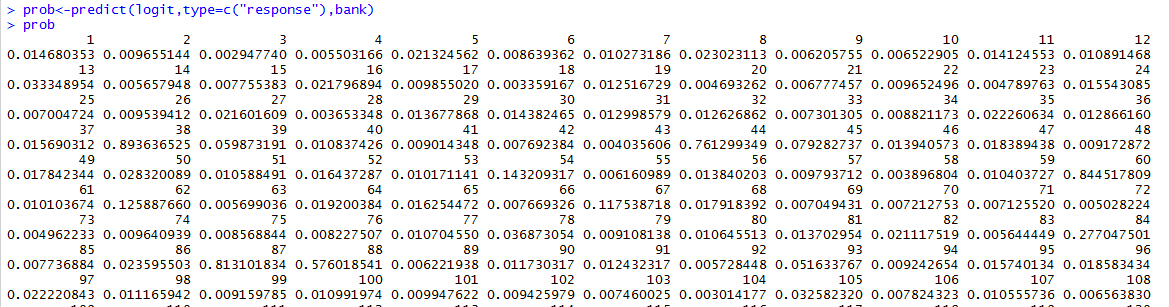
Residual Deviance: 21560 AIC: 21640

1. We shall select the the combination with least AIC value
2. 



1. Since the output is in logarithmic notation, we take the coefficient



1. 
2. We assume the probability greater ha 0.5 to be 1 and less than .5 to be 0
3. To obtain the accuracy we use the following steps :

