STATS 202A – HW6

Logistic Regression :

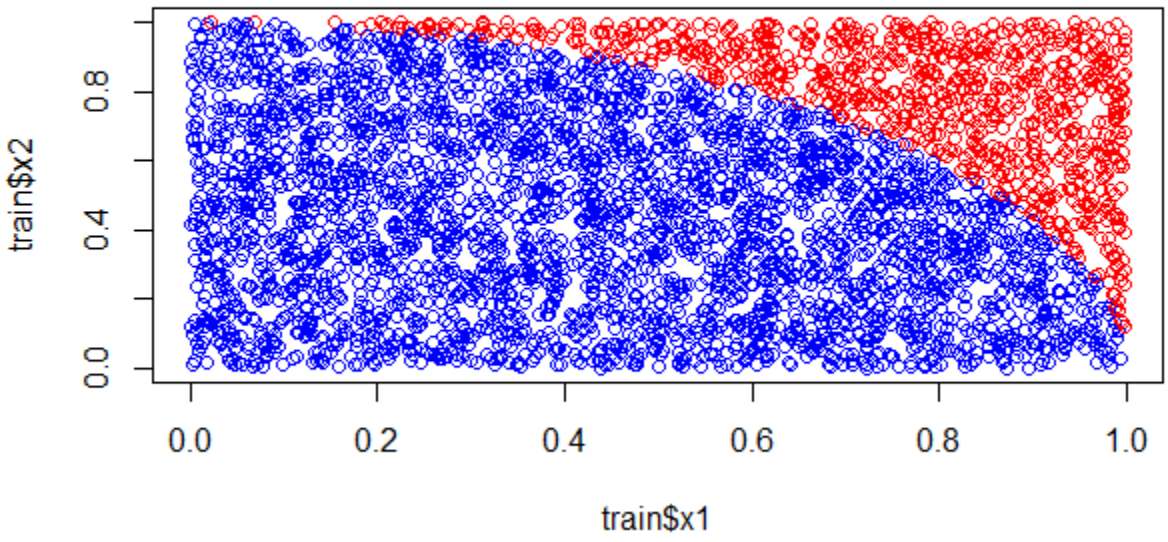
Comparing obtained betas with glm betas

GLM gets much higher accuracy and much lower epsilon so the glm parameters are a little more positive than our positive betas and little more negative than our negative betas.

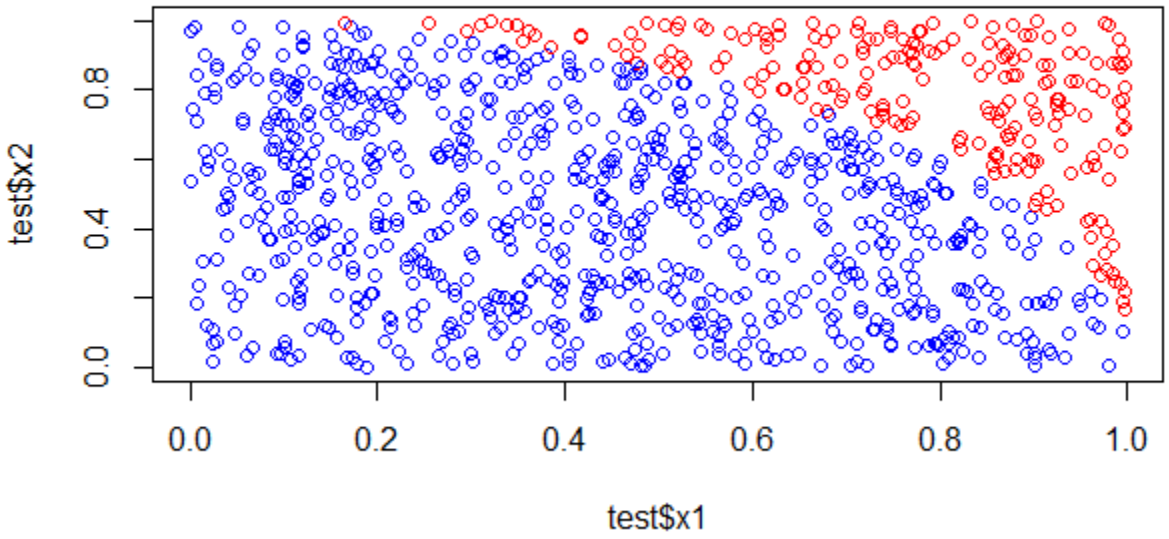
```
          [,1]
[1,] 12.058544
[2,] -2.441365
[3,] -3.053440
[4,]  4.068377
Warning messages:
1: In 2 * as.vector(V %*% t(V))/(t(V) %*% V) :
  Recycling array of length 1 in vector-array arithmetic is deprecated.
  Use c() or as.vector() instead.

2: In 2 * as.vector(V %*% t(V))/(t(V) %*% V) :
  Recycling array of length 1 in vector-array arithmetic is deprecated.
  Use c() or as.vector() instead.

3: In 2 * as.vector(V %*% t(V))/(t(V) %*% V) :
  Recycling array of length 1 in vector-array arithmetic is deprecated.
  Use c() or as.vector() instead.

4: In 2 * as.vector(V %*% t(V))/(t(V) %*% V) :
  Recycling array of length 1 in vector-array arithmetic is deprecated.
  Use c() or as.vector() instead.

> coef(glm(Y ~ X - 1, family = binomial(link = 'logit')))
       X1        X2        X3        X4
14.072039 -2.848363 -3.607550  4.750143
```

ADABOOST:

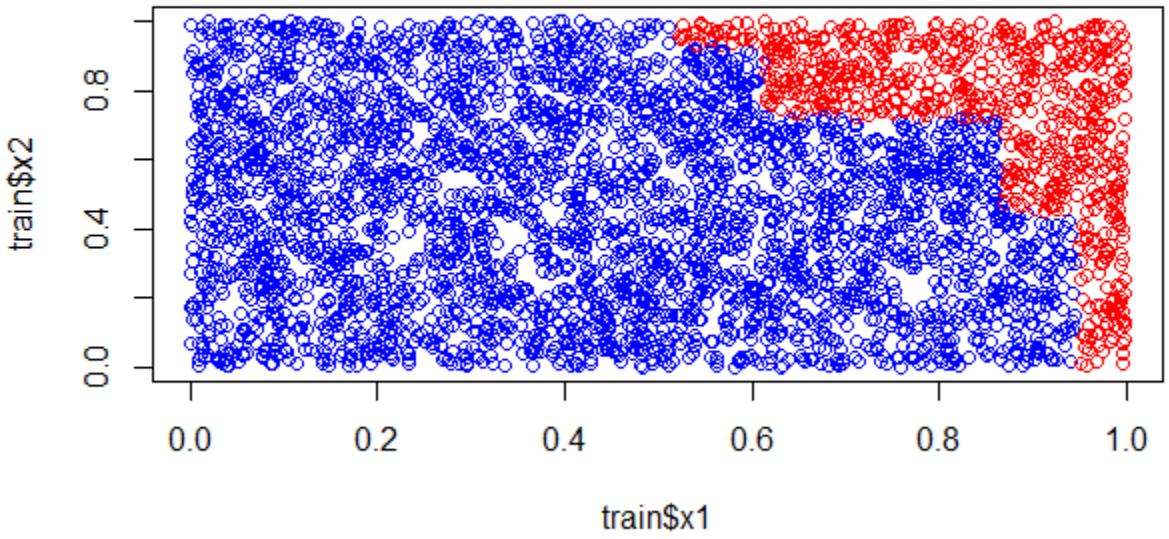
Below is the input training data of 4000 samples with correct y



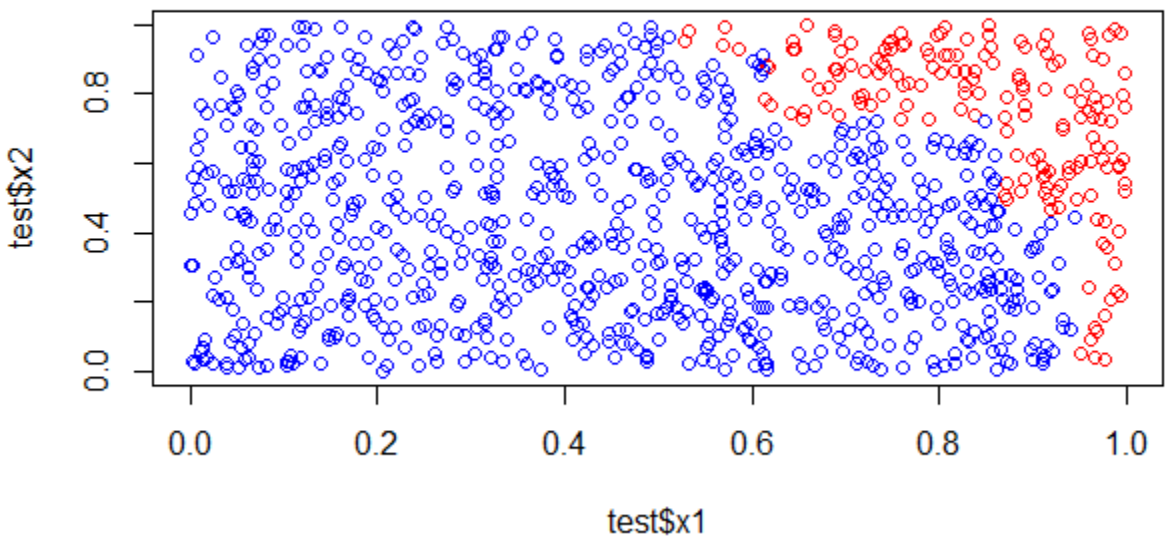
Below is the input testing data of 1000 samples with correct y



Below is obtained for predicted y for training X (4000 samples)



Below is obtained y for predicted y for testing X (1000 samples)



Below is training error and testing error vs no. of decision trees