# Solution to Problem 1

1a)

#defining indicator variable

> IsB <- factor(ALL$BT %in% c("B","B1","B2","B3","B4"))

1b)

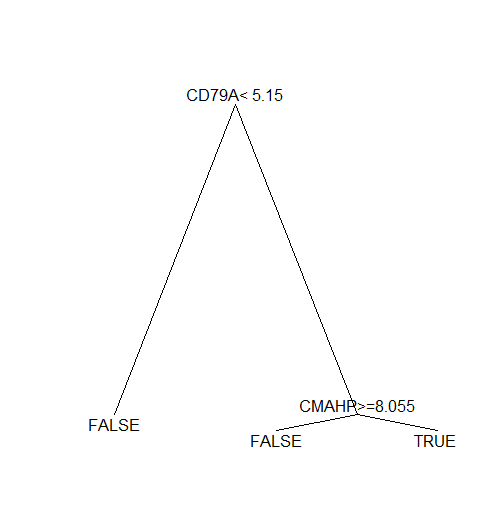
> table(predicted.part, stages)

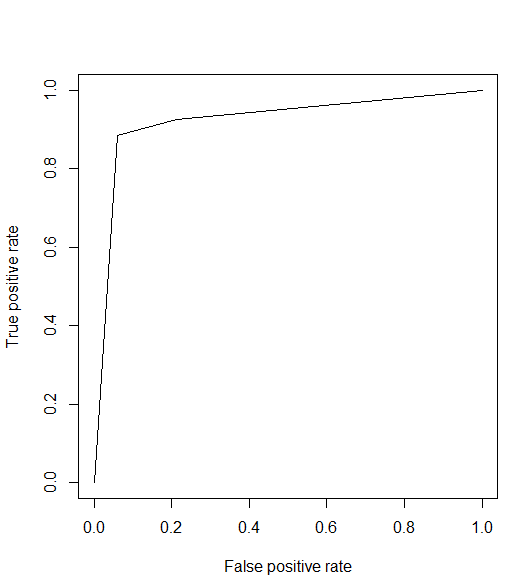
stages

predicted.part FALSE TRUE

FALSE 31 11

TRUE 2 84





1c)

> print("the empirical misclassification rate is")

[1] "the empirical misclassification rate is"

> (11+2)/128

[1] 0.1015625

> # fpr =

> print("the false positive rate is")

[1] "the false positive rate is"

> 2/(31+2)

[1] 0.06060606

> # fnr =

> print("the false negetive rate is")

[1] "the false negetive rate is"

> 11/(84+11)

[1] 0.1157895

> # Specificity = tnr =

> print("the specificity is")

[1] "the specificity is"

> 31/(2+31)

[1] 0.9393939

> print("the area under curve AUC is:")

[1] "the area under curve AUC is:"

> performance(pred,"auc")

Slot "y.values":

[[1]]

[1] 0.922807

1d)

> fnr.true

[1] 0.08310606

The estimate fnr is 0.08310606

1e)

Logistic regression:

> table(pred.B1, IsB1)

IsB1

pred.B1 B not\_B

B 90 6

not\_B 5 27

80% confidence interval for the coefficient of gene "39317\_at"

> confint(fit.data, level=0.8)

Waiting for profiling to be done...

10 % 90 %

(Intercept) -13.767525 -2.8118382

X39317\_at -1.427390 -0.6047588

X38018\_g\_at 2.120174 3.9861802

CI is (-1.427, -0.6047)

1f)