Hints for Assignment 2

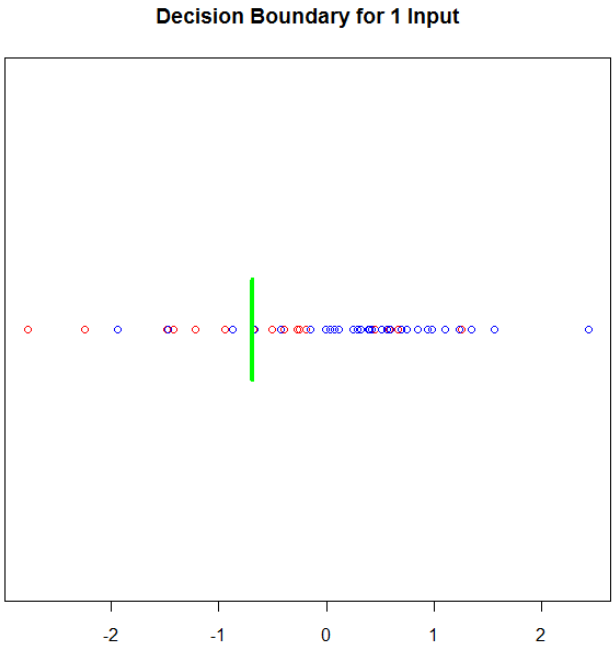
see corrected wording for (a)

Plotting 1D Decision Boundary in (a)

This is just in response to Lihua's question in the lab today.

Using logistic regression, the decision boundary is determined by $\text{logit}(\hat{p}) = \hat{\beta}_0 + \hat{\beta}_1 x = 0$. Solving we see the decision boundary corresponds to a point, $x = -\hat{\beta}_0 / \hat{\beta}_1$. The script below plots the training data and boundary,

```
#Source: DABoundaryOneInput.R
n <- 50 #sample size
b <- 1 #regression coefficient for x1
x0 <- 0.25
x1 <- rnorm(n, 0, 1)
v <- x0 + b*x1
p <- 1/(1+exp(-v))
set.seed(7317)
y <- rbinom(n, size=1, prob=p)
ydf <- data.frame(x0=x0, x1=x1, y=y)
ans<-coef(glm(y ~ x1, family=binomial(link = "logit"), data=ydf))
B<- -ans[1]/ans[2]
Y <- rep(1,n)
plot(x1, Y, yaxt="n", xlab="", ylab="", type="n", ylim=c(0.5, 1.5))
ind <- y==0
points(x1[ind], Y[ind], col="red")
points(x1[!ind], Y[!ind], col="blue")
lines(c(B,B), c(0.9,1.1), lwd=4, col="green")
title(main="Decision Boundary for 1 Input")
```



(d) Singh Microarray Data

Using top 30 genes and row-normalization. Note for the confusion matrix, rows are predictions and columns are observed.

■ linear regression

■ regression confusion matrix for training data

```
> list(CMRegTrain,etaRegTrain)
[[1]]
    y
yfit -1  1
  -1 48  1
  1   4 49

[[2]]
[1] 0.04901961
```

■ regression confusion matrix for test data

```
> list(CMRegTest,etaRegTest)
[[1]]
    yt
yhat -1  1
  -1 24  2
  1   1  7
```

```
[[2]]
[1] 0.0882353
```

kNN

kNN confusion matrix for test data, kNN, k=1

```
> list(CMKNN1Test,etaKNN1Test)
[[1]]
    yt
yhat -1  1
  -1 25  5
  1   0  4

[[2]]
[1] 0.1470588
```

kNN confusion matrix for test data, kNN, k=3

Using the method of Holmes & Adams implemented our R package nnc, k=3 is optimal

```
> list(CMoptTest,etaoptTest)
[[1]]
    yt
yopt -1  1
  -1 25  5
  1   0  4

[[2]]
[1] 0.1470588
```