In-class Assignment

Lab Question

For the following 50 observations (predicted probability and class label), what is the best accuracy value and at what cut-off?

Input is stored in sample1.csv.

- Reading the input file.

```
d <- read.csv("sample1.csv", header = T) #sample.csv
```

- Attach ROCR package

```
library(ROCR)
```

```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'
```

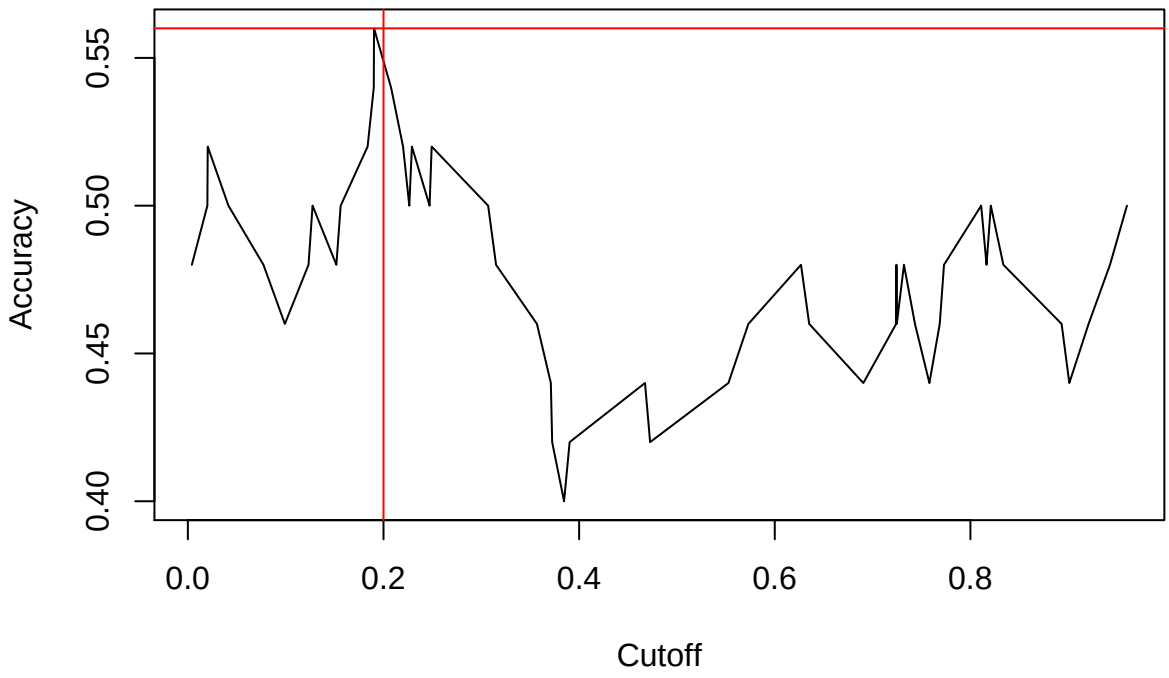
```
## The following object is masked from 'package:stats':
##
##     lowess
```

- Finding the accuracy value and cut-off

```
# Set Column names
colnames(d) <- c("Predicted", "Y")
# Display the dataframe
View(d)

pred <- prediction(d$Predicted, d$Y)
per1 <- performance(pred, "acc")

plot(per1)
abline(h = 0.56, v = 0.2, col = "red")
```



Answer

From the graph we can infer that

Accuraccy = 0.56 at Cutoff value = 0.2

- Verification

```
ifelse(d$Predicted > 0.2, 1, 0) -> py

# Confusion matrix
table(d$Y , py)
```

```
##    py
##      0  1
##   0  8 18
##   1  5 19
```

We can use the confusion matrix, to verify the accuracy.

From above: Accuracy = 27/50 = 0.54 At cutoff 0.2