

# Classification Metrics

*Brett Burk*

*Monday, May 04, 2015*

1 Loading in the data:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
folder <- "C:\\Users\\Brett\\Dropbox\\CUNY\\621\\Classification Metrics\\"
pregnancy <- read.csv(file.path(folder, "classification-output.csv"))
```

Get the total amount of positive values

```
baseline.event <- sum(pregnancy$class)
```

Order the pregnancy data by the probabilities

```
pregnancy <- pregnancy[order(pregnancy$Scored.Probabilities, decreasing=TRUE),]
```

Gives a vector of the correct.classification percentages

```
correct.classifications <- pregnancy$class/baseline.event
cumul.classifications <- cumsum(correct.classifications)
```

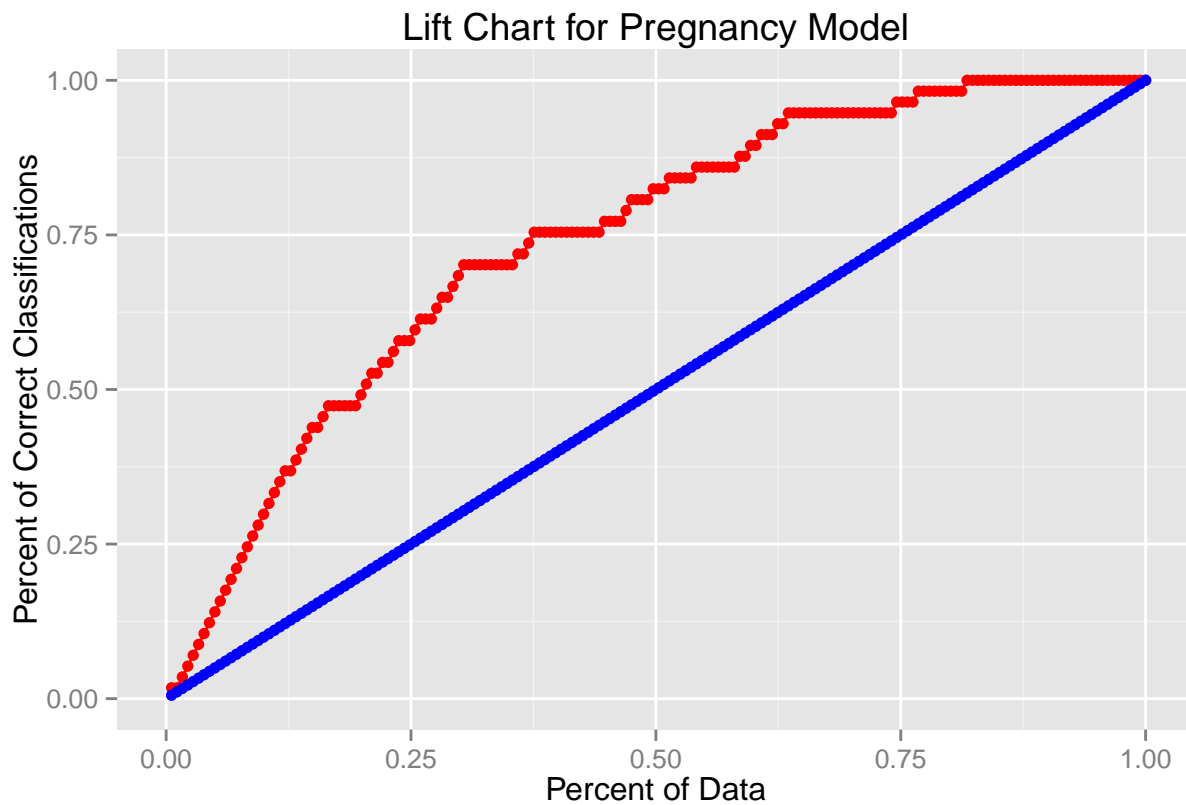
Generic line

```
to.combine <- (1:(length(correct.classifications)))/length(correct.classifications)
baseline <- data.frame(cbind(to.combine, to.combine))
names(baseline) <- c("x", "y")
lift.data <- data.frame(cbind(to.combine, cumul.classifications))
```

The Plot

```
p <- ggplot() +
  geom_point(data = lift.data, aes(to.combine, cumul.classifications), color="red") +
  geom_line(data = lift.data, aes(to.combine, cumul.classifications), color="red") +
  geom_point(data = baseline, aes(x, y), color="blue") +
  geom_line(data = baseline, aes(x, y), color="blue") +
  ylim(0,1) +
  xlim(0,1) +
  xlab("Percent of Data") +
  ylab("Percent of Correct Classifications") +
  ggtitle("Lift Chart for Pregnancy Model")
```

p

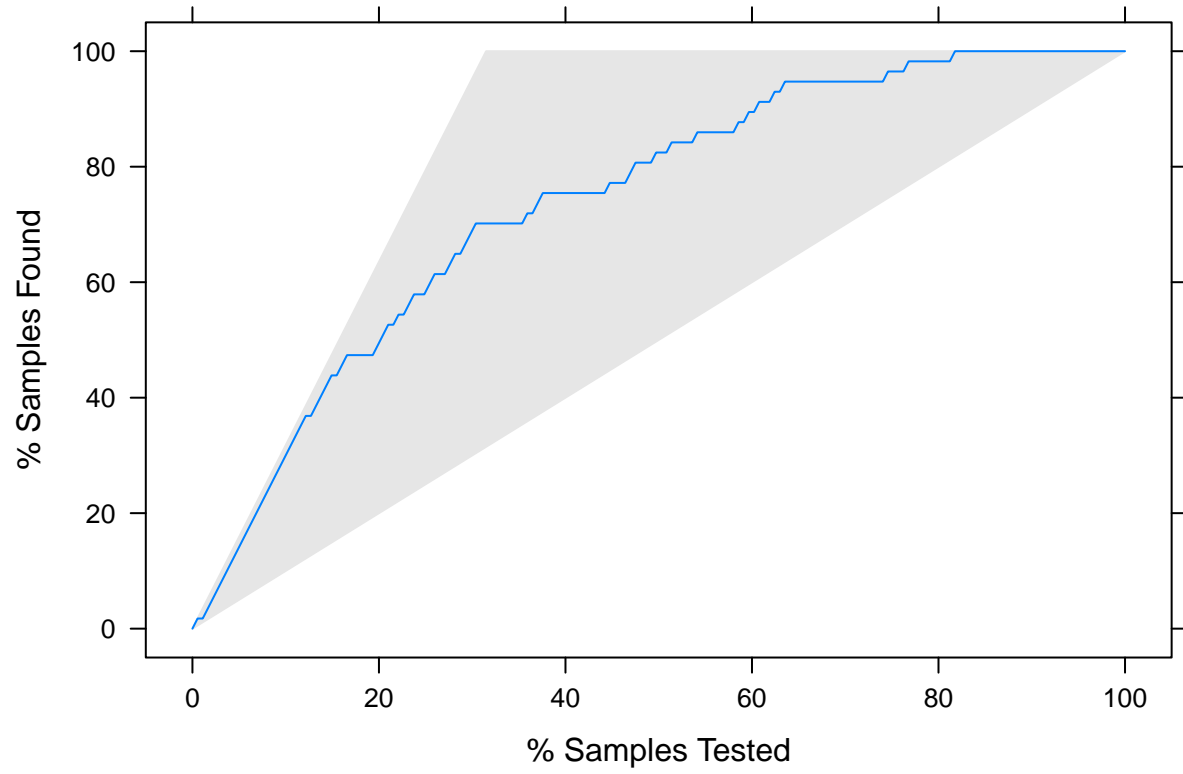


```
library(caret)
```

2

```
## Warning: package 'caret' was built under R version 3.1.3
```

```
# This was a little annoying to figure out, until I realized that it confused my positive and negative  
lift1 <- lift(as.factor(1-pregnancy[,9]) ~ pregnancy[,11], data = data.frame(pregnancy[,10]))  
xyplot(lift1, plot = "gain")
```



Both models appear the same, although the lift function in caret is obviously a lot simpler to operate.