ConfusionM

Nikhil Muthukrishnan

6 February 2019

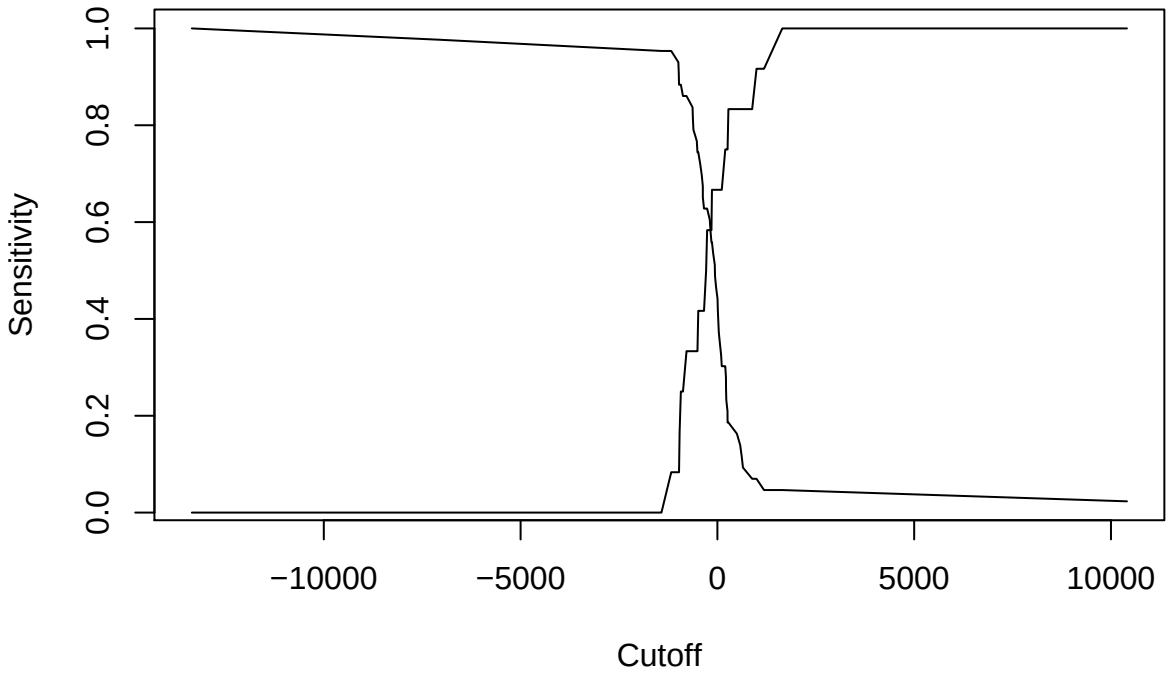
Objective:Build model to predict if batch is on/off spec based on sensor means Contents:Head of sample(2017,all.pv,L2) +model call +threshhold value set by specificity and Sensitivity. +Confusion Matrix +Specificity and sensitivity

Criticisms:Assuming mean values will be good representativ +Assuming new 5 minutes data is aggregated from poin to chang +Class imbalance is still pervasive

```
##     Inspec SC2_PIC20002.pv SC2_LIC20064.pv SC2_FIC20464.pv SC2_FIC20463.pv
## 1: OFFSPEC        36.67230        64.15678        1.681969        1.660420
## 2: OFFSPEC        39.97620        60.16522        2.917113        2.883212
## 3:  ONSPEC        39.99079        60.04478        2.496099        2.492556
## 4:  ONSPEC        39.99560        60.09578        2.932488        2.909104
## 5:  ONSPEC        40.00669        60.15119        3.219294        3.209181
## 6:  ONSPEC        40.00689        60.00831        3.179044        3.152190
```

Model used for training has been oversampled,undersampled and is now in balance with equal numbner of on and offspec batches.

model <- glm(as.factor(Inspec)~., data = balanced, family = binomial(link="logit"))



Logistic regression gives results as Log odds. The predicted values have to be given a threshold value to round the probabilities as classes.

The ideal threshhold value is where the line for specificity and sensitivity intersect, ie.0-0.5

```
##            Reference
## Prediction OFFSPEC ONSPEC
##    OFFSPEC       8     24
##    ONSPEC        4     19
```

```
##          Sensitivity          Specificity       Pos Pred Value
##            0.6666667            0.4418605            0.2500000
##       Neg Pred Value            Precision               Recall
##            0.8260870            0.2500000            0.6666667
##                   F1           Prevalence       Detection Rate
##            0.3636364            0.2181818            0.1454545
## Detection Prevalence    Balanced Accuracy
##            0.5818182            0.5542636
```

The Confusion Matrix shows that out of 55 total observation in the test set offspecs were 12(7+5) and onspecs were (35)14+29, of which 7/12 offspecs were correctly predicted and 29/43 onspec batches were also predicted.