Demo

Shukry Zablah

05 December, 2018

Contents

Imports

```
#import libraries
```

Load Model

```
clf <- readRDS(file = "../models/LogisticRegressionClassifier_Full.Rds")
```

Load Test Set

```
test <- readRDS(file = "../data/PIMA_test.Rds")
```

Predicting Cases

```
woman5 <- test %>%
  filter(row_number() == 5)

woman18 <- test %>%
  filter(row_number() == 18)

woman37 <- test %>%
  filter(row_number() == 37)
```

We can see the single woman's observation from the test set. Now we predict what's the probability she has diabetes.

```
predict(clf, type = "response", newdata = woman5)
```

```
##         1
## 0.7885501
```

And in fact we check and woman5 does have diabetes.

```
predict(clf, type = "response", newdata = woman18)
```

```
##         1
## 0.1652994
```

And once again our model was correct, woman18 does not have diabetes.

```
predict(clf, type = "response", newdata = woman37)
```

```
##         1
## 0.2049365
```

Our model is off here. It gives us a low probability prediction but the woman does indeed have diabetes. This is the kind of examples that we wanted to minimize by choosing a 0.3 cutoff. However, a lower cutoff would have been needed for this woman (at 0.2) to classify her correctly.