Siddharth Patel
CS 484 – 001
Date: 11/7/2021

Overview

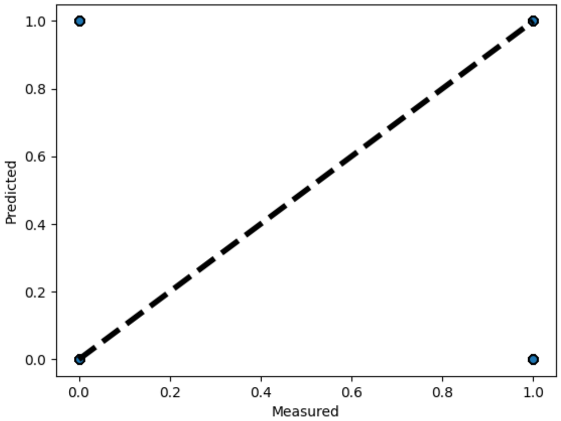
Up until now, we have been implementing cross validation to check our algorithm accuracy in Data Mining, but we decided to move a step further by implementing five-fold cross validation in this assignment. What is five-fold cross validation? This procedure has a single parameter called k = 5 that refers to the number of groups that a given data is to be split up into. As such, the procedure is often called five-fold cross validation. Cross validation is primarily used in machine learning to test/ estimate how a model is expected to perform in general when used to make predictions on data not used during the training of the model. The first step is to shuffle the dataset randomly (optional), then split the dataset into five groups, and process on each dataset however you choose.

To measure the bias of the dataset, concepts such as true positive, false positive, true negative, and false negative are invented. A true positive is an outcome where the model correctly predicts the positive class. A true negative, on the other hand, is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class. A false negative, however, is an outcome where the model incorrectly predicts the negative class.

Analysis

The first task for this project is to implement the five-fold cross validation to get either 0 or 1 prediction for every individual in the training set. Initially, I am using the “race” variable as a feature, but I do remove it later. To implement the five-fold cross validation, I am using the cross_val_predict().

Once I implement the five-fold cross validation followed by getting the 0 or 1 prediction for every individual training set, I make sure to graph the predictions.

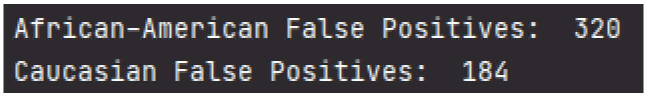


As you can see on your right, my prediction represents the four primary building blocks of the metrics we use to evaluate classification models: false and true positives, and false and true negatives. Since our output can only be 0 or 1, we have 4 dots in each corner based on (0, 0), (0, 1), (1, 0), and (1, 1).

Approaches

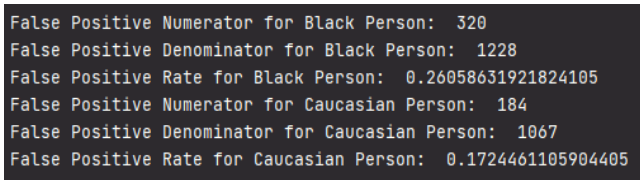
To accomplish my first question, I went ahead and filtered each row such that for data1, the race would be African Americans who did not actually recidivate. For data2, the race would be Caucasians who did not actually recidivate. Note that I follow a similar procedure for the second question. Once the data was separated, my goal was to figure out whether this machine learning model is biased towards black men.

I further computed the false positives of African Americans followed by Caucasian using the support vector machine model that I used in previous project. The output is shown as reference on the right.



```
African-American False Positives:  320
Caucasian False Positives:  184
```

After computing the false positives, I proceed by calculating the false positive numerators, denominators, and rates (FPR) for each race which resulted in the output to the right.

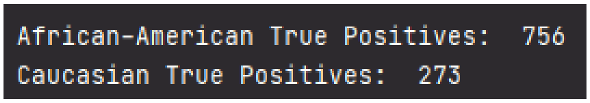


```
False Positive Numerator for Black Person:  320
False Positive Denominator for Black Person:  1228
False Positive Rate for Black Person:  0.26058631921824105
False Positive Numerator for Caucasian Person:  184
False Positive Denominator for Caucasian Person:  1067
False Positive Rate for Caucasian Person:  0.1724461105904405
```

As you can see, the likelihood of a black individual to be falsely arrested and charged with another crime in the following two years is higher compared to a Caucasian individual since the difference between the false positive rates between the two races is almost 10%. Hence, this proves that the SVM model that I used for prediction is biased AGAINST African Americans. The implications this can have on our society has already been witnessed in the recent Black Lives Matter moment where the blacks were considered to be the race with high likelihood to commit crime. It also gave Caucasian men a sense of superiority due to which many felt comfortable harassing or inflicting physical wounds on Black individuals. Caucasian polices officers murdered a black men who clearly was not guilty and was constantly pleading for mercy which the officer was choking him with his knee. This superiority corrupted the legal system which led to a moment with the hopes of justice to the innocents of the past.

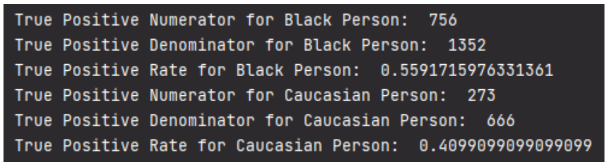
For the second question, I had to implement model calibration prior to any computing. Note that model calibration is the process of adjustment of the model parameters and forcing within the margins of the uncertainties to obtain a model representation of the process of interest that satisfies pre-agreed criteria. This resulted in another name for it “Goodness of Fit”.

I used the Calibrated Classifier to calibrate the Support Vector Model that I had. My further process looks identical to the previous question. However, the output was drastically different when I ran my code. Let’s go ahead and analyze.



```
African-American True Positives:  756
Caucasian True Positives:  273
```

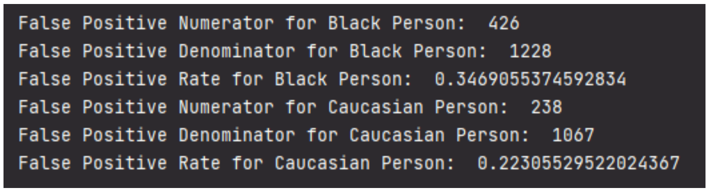
Here is the calibrated true positive computation related to the two races. As you can see, more black individuals are expected to be rightfully charged or arrested with another crime in the following two years. The increase in likelihood for black men to be more likely be guilty is due to the calibration.



```
True Positive Numerator for Black Person:  756
True Positive Denominator for Black Person:  1352
True Positive Rate for Black Person:  0.5591715976331361
True Positive Numerator for Caucasian Person:  273
True Positive Denominator for Caucasian Person:  666
True Positive Rate for Caucasian Person:  0.4099099099099099
```

To further assess the results of this calibration, I went ahead and computed the True Positive Rates (TPR) along with the numerator and denominators of the two races.

In my conclusion there is an even greater bias claiming that Black men are more likely to be rightfully arrested and charged with another crime in the following two years compared to a Caucasian individual since the difference in the probabilities between the two is even higher; it reaches almost 15% (clearly higher than the previous 10%). Due to this enormous bias, many people fear approaching or interacting with Black men thinking they are scary which is not the case in most scenarios. As a result, they are considered the outcasts of the society in many people’s eyes. On the other hand, Caucasian race is given more freedom due to their lighter skin. The implications of this statistics are not just on black community, but any community with dark skin tone. This has been a result of racism since a long time.



```
False Positive Numerator for Black Person:  426
False Positive Denominator for Black Person:  1228
False Positive Rate for Black Person:  0.3469055374592834
False Positive Numerator for Caucasian Person:  238
False Positive Denominator for Caucasian Person:  1067
False Positive Rate for Caucasian Person:  0.22305529522024367
```

I think that the metrics without calibration is more appropriate in this domain. My thought process is to eliminate bias into consideration when I am choosing the two metrics because both will come with their bits and pieces of biases. If you look at the first metric, you can see that false positive rates are lower. To back up my statement, I have attached false positive rates with calibration. As you can see, the false positive rates with calibration has higher chance of falsely accusing individuals no matter which race. On the other hand, the false positive rate without calibration will be less likely to falsely accuse people. Moreover, the True Positive rate is also lower without calibration whereas the True Positive rate with calibration goes all the way up to 50%. This means that 1 out of every 2 individuals will be considered rightfully accused based on the model.

By choosing the first model (one without calibration), I can lower the rate of false accusations and the amount of crimes or arrests that will take place in the future.This will hopefully give rise to a more secure society where everyone can live in peace and harmony. Ofcourse, my hypothesis is based on the conclusion that the models will be accurate.

Accuracy Of SVM: 0.6264026402640264   Accuracy w/o Race Feature: 0.6248762131115072

There is little to no change in the accuracy with or without the race feature. This gives me a reassurance in a way that even if a particular race is not considered in the accuracy of my model, I will still have a non/ less biased model that will be fair in judging individuals and punishing them for their deeds. Even if you remove the protected features, it will not drastically impact the accuracy of the algorithm ensuring a steady output.

Conclusion

Although, in both my findings, I concluded that my models were biased, I must admit that to get to this conclusion with just simple statistics is a bit unfair since a lot of factors are not considered when computing the calculations for these biases. A person who is walking around town may have committed multiple crimes but left no traces to be caught. Hence, proving himself not guilty.

To measure the accuracy of the model, we still have not looked at crimes in foreign countries such as China, Japan, India, Australia, United Kingdom, and many more. Will it result in the same output? That is another domain this model has yet to explore.