MACHINE LEARNING HW 1

Q1: Suppose a crime has been committed. Blood is found at the scene for which there is no innocent explanation. It is of a type that is present in 1% of the population. A suspect who has this rare blood type has been charged with the crime.

The prosecutor claims: "There is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus, there is a 99% chance that he is guilty." This is known as the *prosecutor's fallacy*. What is wrong with this argument?

Answer: According to Bayes Theorem the relation between conditional probability and its inverse is P(A/B) = (P(B/A) * P(A)) / P(B)

In the question it was stated that P(murder / person with rare blood group type) = P(person with rare blood group type / murder) which is wrong according to Bayesian Model. P(A/B) != P(B/A). for example P(Apple / Fruit) != P(Fruit / Apple).

Let's examine this problem in detail, let's assume the population of the city is 10,000 people and 1% of the population (100) have a rare blood group type.

P(Match/ Innocent) = 1% i.e.. 1/10,000

Guilty/ Innocent	Match	No-Match
Guilty	1	0
Innocent	99	9,999

P(Guilty/ Match) = 1% in 1% of total population i.e.. 1 / 99+1 = 1/100

Naming convention followed in Guilty row:

1 → the person who has the rare blood group type and is guilty

0 → the person who doesn't have the rare blood group type and isn't guilty

The question states that "1% of the population has rare blood group type". in the above example we have 100 people with rare blood group type. Now, the prosecutor picked 1 person out of the 100 people and claimed that he/she is guilty. The probability of picking guilty out of the 100 people is 1/100 which is very rare. hence the prosecutors claim is wrong. Among 100 people only one person will commit the crime, not all the 100 people with rare blood group type will commit the crime. This can stand as an induction proof of Bayes theorem.

Q3: What is your assessment of criminal risk scores? What should be their role in the criminal justice system? What metric should they optimize for: False Positive, False Negatives, True Negatives, True Positives, Precision, Recall, F-Score, others? Provide an argument and example for your thesis.

Answer: Criminal Risk Score reports the likeliness of a defendant committing a new crime. If the score is low then the chance of him committing a crime again is less , if the score is high then the chance of the defendant committing a new crime is very high.

These Criminal score reports are provided to the Judges as an insight on the defendants. But there is a chance that the model can produce biased results.

Below is an example of the results produced by the model: (screenshot from Ethics in ML PPT)

From the case study conducted by ProPublica, they collected criminal records of 7000 people from Florida who committed crimes in 2013-2014. The thesis proved that the model was labelling black people with low criminal record as high risk criminals and , white people with high risk background as low risk criminals which proves that the model is biased.

The judgements made by judge can be biased towards a humans race, gender, colour etc.. this data should not be given as training data to the model, if given then the model will produce biased results which are no different than human made decisions.

Some metrics which needs to be optimized:

The confusion matrix results should be optimized, specifically **False positive is one metric which needs to be reduced**. Because True Positives and True Negatives are the correct results produced by the ML model. False Negative results can be optimized later but **False positive** will be one feature which will provide injustice to the defendant. If a model is producing high False Positive rate then the model needs to be revised and checked if it is biased. False Positive conveys that the person who has not committed crime is convicted for no mistake done by him. By reducing False Positives we can provide justice to people.

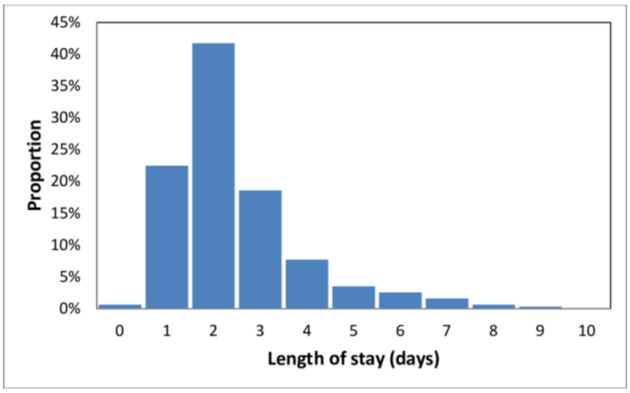
Precision: precision = True Positives / (True Positives + False Positives)

Again the False positive metric in Precision needs to be as less as possible to produce the most accurate results. If the False positives are 0, then the ML model has produced 100% correct result.

Recall: recall = True Positive / (True Positive + False Negative).

In this case we need to optimize for False Negatives because , if the FN rate increases then the Recall value will be disrupted. False negatives means, that the person who committed the crime has not received any punishment , this will not provide justice to criminals. By reducing the False Negative rate we can optimize for Recall value.

Q4: Suppose the distribution of length of stay of patients in a hospital is given by the graph below. You are given the task to predict the length of stay. How would you change the problem from a regression problem to a classification problem?



Answer: This regression problem can be converted into classification problem by using a technique called Discrete Binning. The target variable can be divided into classes by splitting them into n intervals.

Discrete Binning : Binning can be done in 2 ways, Equal Width Binning and Equal Frequency Binning. For this example I will be using Equal Width Binning, I will divide the data according to the following formula . Width = (max – min) / 3 i.e.. into 4 bins each of size 3. So, the bins would look something like this: Bin1 [0 – 2] , Bin2 [3 – 5] , Bin3 [6 – 8], Bin4 [9 – 10] .

We can also apply Equal Frequency binning process, where we will be diving the data into k groups where each group contains approximately equal number of records.

Q5: What should be the role of machine learning models be in making hiring decisions? Should there be legislation to limit the proliferation of such models, or should they be regulated in some other way since they can be made to be less biased as compared to humans?

Answer: it was mentioned in the class that Amazon has used an ML model in its recruiting process and had massive bias towards women. The training data provided to the model was about historical data of successful candidates, it might be possible that all the candidates who were successful might be men, which is when model started building an assumption that only men can be potential candidates and became biased towards women. Due to lack of correct data the model became biased and started making wrong decisions.

There should not be any legislation to limit the proliferation of models, instead they need to be regulated to be less biased. ML model plays a major role in saving hiring managers time by screening resumes and finding out the potential candidates. Some of the regulations that can be applied to existing models are by making the model to consider features which are important to the job requirement ex: work experience, skill, previous projects, extra curriculars etc, we should also train the model to not consider features like Gender, race, neighbourhood etc. By placing conditions on what features that the model shouldn’t consider, the current performance of the ML model can be significantly improved.

(Extra Credits).

Q: Why is the performance of a prediction algorithm not guaranteed to be the same in production as compared to its performance on test set in the original dataset?

Answer: the performance of the prediction algorithm is not guaranteed to be same in production because the data used for training the model before production is historical data that is correct and not skewed, by testing the model with proper data it will be producing best results with high accuracy. But when the model is deployed in production it needs to learn from the current data and needs to produce an output by combining current data with historical data. The current data which is being fed to ML model might be skewed and the model might mis interpret the data. The model will then produce wrong results. We need to build model in such a way that it needs to understand which features to consider and which to ignore. By doing so we can optimize for False Positives and False Negatives.

Assignment by: Venkata Ramani Srilekha Bandaru