CS 3600 Project 4 Wrapper

CS 3600 - Fall 2023

Due December 8th 2023 at 11:59pm EST via Gradescope

Introduction

This Project Wrapper consists of a context paragraph, which identifies the topic of the wrapper, followed by four short-answer questions, each worth 1 point. Please limit your response to each question to **a maximum of 200 words**. The goal of this assignment is to train your ability to reason through the consequences and ethical implications of computational intelligence.

Context

You are working for a financial institution which is in the business of assessing risk of loaning money to potential clients. The CTO of the company has been an avid Twitter user swept up in the rumor that GPT-4 will have 100 trillion parameters, and has a dream to completely automate the loan risk assessment process. As a former CS3600 student and experienced engineer in the AI/ML team for the company the CTO comes to you to ask for a plan to implement this algorithm. You have historical data since the 1930s in this format (Note that all strings come from a fixed, enumerated set of values):

Name	**Pronouns**	**Sex**	**Age**	**Zip Code**	**Race**	**Hobbies**	**Credit Score**	**Net Worth**	**Loan Amount**	**Accepted for Loan (target)**
string	string	string	int	int	string	string	int	int	int	boolean

Table 1: Loan Dataset

Question 1

Fully engaged in the Twitter hype for GPT-4 and constantly up to date on the NeurIPS threads the CTO pitches a neural network architecture for predicting if the client will default on the loan. The CTO suggests the various columns of the table act as features to a deep neural network with 50 hidden layers. With your experience in the field you think that there could be a better model for such a sensitive task as this. Make an argument for why a decision tree is better for use in this particular situation (hint: you can mention runtime at inference, but there is an even stronger case to be made for a decision tree that we are looking for here! Especially consider that this ML model will be making highly sensitive decisions.)

Answer: Decision trees offer a better solution in this scenario for several reasons. Firstly, decision trees are more interpretable than deep neural networks. This interpretability is crucial when making sensitive decisions like loan approvals, as it allows for easier identification and correction of biases in the model. Secondly, decision trees handle feature interactions more naturally, which is important given the diverse range of features (like age, race, net worth) in the loan dataset. Lastly, decision trees have a faster inference time compared to a deep neural network with 50 hidden layers, making them more efficient for real-time decision making.

Question 2

The ethics review board at the company rejects the initial proposal from the CTO on the basis that algorithm could easily end up rejecting loan applicants based on race or sex. To fix this the CTO proposes that the sex and race columns be removed from the dataset. Will this completely prevent the machine learning model from discriminating based on race/sex? Why or why not?

Answer: Simply removing the sex and race columns from the dataset does not completely prevent discrimination. This is because other features in the dataset might correlate with race or sex, leading to indirect discrimination. For instance, zip codes or hobbies might have associations with specific racial or gender groups. Thus, the model could still learn to discriminate based on these proxy features. It's important to not only remove direct identifiers like race and sex but also to monitor and adjust for indirect biases that can arise from correlated features.

Question 3

Algorithmic discrimination is a serious problem with the wide dissemination of machine learning models [**10.5555/3002861**]. Unfortunately, these harms can go unnoticed due to the false assumption that math and algorithms are unbiased and objective. Machine learning models are only as good as the data used to train them. Research an instance where a machine learning model was used to make critical decisions, but was later found to be biased (excluding the example with bank loans). Summarize what the purpose of the model was, and how it ended up causing harm. Include a link to a news article or research paper discussing this issue.

Answer 3: A significant instance of machine learning bias was observed in facial recognition technologies used by various law enforcement agencies. Studies, including one by the MIT Media Lab, revealed that these systems exhibited higher error rates in identifying women and people of color compared to white men. The root cause was traced back to the training datasets, which were predominantly composed of images of white males. This bias led to wrongful identifications and arrests, raising serious ethical concerns about the deployment of such technology in law enforcement without thorough bias mitigation. The case highlights the critical need for diverse and representative training data in machine learning models to prevent discriminatory outcomes. (Source)

Question 4

What is the reason for having a separate train and test set when constructing a machine learning model? Why not use all the possible data for training and testing?

Answer: Separating data into training and test sets is crucial in machine learning to evaluate the model's performance on unseen data. The training set is used to fit the model, while the test set is used to assess its generalization capability. Using all data for both training and testing can lead to overfitting, where the model performs well on the training data but poorly on new, unseen data. The separation helps ensure that the model is robust and can make accurate predictions in real-world scenarios, beyond the examples it was trained on.