**Eth. AI, HW2, Bias**

**Due Thursday, Week 5, 11:59PM**

**Goal:** Find bias in models and mitigate it.

**Resources:** Tutorials on python and sklearn library can be found both in [W3Schools](https://www.w3schools.com/python/default.asp) as well as [here (Python course Slides](https://ind657-my.sharepoint.com/:f:/g/personal/jrusert_pfw_edu/Ei_EBaJU-IJBgpwM2A7mmvUB-GrO9jUSvSq0CgMgwZgwYw?e=DJcAgX)). Also, sklearn has demonstrations of models and functions on the specific model pages as well (see below links).

**Models:** You will be working with 1 type of model for this homework, from the sklearn python library:

[Logistic Regression](https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.LogisticRegression.html)

This homework consists of 2 parts. Even though there are 2 parts, there still should only be 1 final report and 1 .py file handed in.

## Case Study 1

Your boss has exciting news to share with you. He has used AI to generate code which helped him train a model to automatically approve or deny loan requests. The exciting part is it achieved perfect (100%) accuracy on a test set of data he curated from existing loan data. He has saved the model via pickle and shared it with you.

You test the model on a held out set of loan data and find it to perform poorly (~47% accuracy). You believe the issue may be due to bias in the model. It is your goal to find out the bias and present it to your boss.

**Coding goals:**

* Load in the trained model (BEST\_LR\_MODEL\_EVER.pkl) and verify the above claims first using pickle (an example can be seen below).

import pickle

import pandas as pd

from sklearn.linear\_model import LogisticRegression

from sklearn import preprocessing

with open('BEST\_LR\_MODEL\_EVER.pkl', 'rb') as f:

clf = pickle.load(f)

df = pd.read\_csv('test\_bias\_new.csv')

X = df.loc[:, df.columns != 'approved']

Y = df['approved']

norm\_X = preprocessing.normalize(X, axis=0)

print("New Test Score:", clf.score(norm\_X, Y))

* Locate the bias in the model.
  + Your boss **does not understand** weights and coefficients so this will not be adequate.
  + Your boss **does understand** confusion matrices, so you should leverage these to present a clear picture of bias.
* Hints:
  + Since the new test set shows the poorest performance, this will be the one you’ll want to examine.
  + To normalize the data, your boss used: preprocessing.normalize(X, axis=0)
  + You can send in single queries to models using the .predict function which could help you test bias theories. (Note single queries are not adequate demonstrations of bias for your boss).
  + Once you locate the bias, you can use subsets of the new test set to better demonstrate the bias (via confusion matrices).

**To be added to the report:**

* Describe the steps you took to find the bias. (You should include ones that didn’t help you find the bias as well, to show your thought process.)
* Say what the bias was and why it was affecting the model.
* Demonstrate how you know the bias is there (e.g. confusion matrices, graphs)

## Case Study 2

Your professor tasks you and a friend with creating a simple model to act as an early warning system and predict whether a student will pass a class based on three things: average homework grade so far, exam 1 score, and attendance percentage. You are provided with a training set of 1600 and a test set of 400 examples. Your friend trains a logistic regression model and achieves accuracies of 69 and 71 for the train and test sets respectively. She believes that without more data, the project is hopeless. You have a hunch that some bias in the dataset may be affecting the results. Your goal is to examine the dataset, find the bias, and mitigate it to improve the LR model.

**Coding goals:**

* Load in the trained model (**PassPredictionLR.pkl**) and verify the above claims first using pickle (an example can be seen in the previous study).
* Locate the bias (e.g. the reason the model is performing poorly).
  + Hints
    - the bias is a different type than the first study.
    - Confusion matrices can be helpful here again.
* Fix/mitigate the bias.
  + Hints
    - The technique can be simple, make sure you first know where the bias is coming from before trying to fix it.
    - There are simple techniques shown in the slides. Discuss with the professor if you are lost.
    - The mitigation should improve the test accuracy. You should not modify the test dataset or the results will be invalid.
    - Your coworker also used preprocessing.normalize(X, axis=0)

**To be added to the report:**

* Describe the steps you took to find the bias. (You should include ones that didn’t help you find the bias as well, to show your thought process.)
* Say what the bias was and why it was affecting the model. Demonstrate how you know the bias is there (e.g. confusion matrix, graph)
* Describe the steps you took to mitigate the bias. (Again, include the ones that didn’t help.) Demonstrate that you addressed the bias (e.g. confusion matrix).
* Include a table that includes the accuracies (test and train) before and after the mitigation, to show the improvement.

**To Turn In:**

* Python file (**USERNAME\_HW2.py**.) which contains functions related to the above studies. You should include comments and useful function names to differentiate the two case studies.
* Report which contains:
  + The above requirements for each study. Make sure you make it clear (via titles or subheadings) which case study you are describing.
  + Discussion of issues or difficulties encountered.
  + A note of any AI tools used in coding or writing.

**Additional Rules (MUST BE FOLLOWED):**

1. All Homeworks should follow the overall [homework guidelines](https://ind657-my.sharepoint.com/:p:/g/personal/jrusert_pfw_edu/EZxf1ZsRXjBEkQLNhEhdTgUBt6U64KiT1DJ1YHtkARgKwA?e=bwHrbf)
2. The code should be written in python 3.
3. If noted, the functions must follow the naming and number of arguments as demonstrated.
4. You should make your code modular to the different steps. (You may have more functions to help your main functions)
5. You should be adding comments to document your code. **If I can’t understand why you perform an action, then I can’t credit you for performing that action.**
6. The report should be readable and reference your code, **without explicitly including code.**
7. You should include your name and homework number in the comments at the beginning of the python file.

**Report**

The reports for the homeworks are necessary to communicate your learning and thinking through of the material. Examples of good reports can be found on brightspace under Additional Resources/Guides. Note that your report style may differ, but it is a good reference to start with.

**Grading**

Assignment will be graded as follows:

|  |  |
| --- | --- |
| **Description** | **Points** |
| **(.py file)** Case study 1 coding requirements | 10 |
| **(.py file)** Case study 2 coding requirements | 10 |
| **(Report)** Case study 1 report requirements | 10 |
| **(Report)** Case study 2 report requirements | 10 |
| **(Report)** Other Report Requirements | 5 |
| **(.py file)** Documentation/Code (Comments, functions, etc) | 5 |
| **Total:** | **50** |

* **If the code does not run, it cannot be graded well.** (Many points can be lost if the code cannot be run, as I will not be able to fully test the implementation of the functions).
* **Breaking of the additional rules can result in applied penalties.** (Always make sure you are checking against the rules)

**Suggestions**

* **Documentation is key for showing your effort in this homework.** Make sure you are noting why you make certain decisions all throughout your code.
* The slides for previous classes are posted, so please refer to these and the book for ideas during implementation.
* Start simple, build up complexity. You should always make sure your new ideas being added do not cause your program to crash. So starting simple is the best way to a) maintain the ability to keep your code running, b) add in comments for documentation and thought process as you add more code.
* Work through the homework yourself, rather than sharing ideas (especially not code) with other students. **As a reminder, plagiarism (or sharing) of code is strictly prohibited.** This assignment is complex enough that significant overlap between students will be suspicious.
* If you have not worked with python before, w3schools can help you translate your previous coding experience to python (<https://www.w3schools.com/python/default.asp>)
* Stop by office hours to discuss ideas. I am always happy to help you think through your process!