ENGR 421/DASC 521: Introduction to Machine Learning
Homework 6: Area Under the ROC Curve
Deadline: June 4, 2025, 11:59 PM

In this homework, you will implement the receiver operating characteristics (ROC) curve in Python. Here are the steps you need to follow:

1. Read Section 20.7 from the textbook.

2. You are given the true labels of 1000 data points for a binary classification problem in the file named `hw06_true_labels.csv`, the predicted posterior probabilities of these 1000 data points for the positive class obtained using `Algorithm 1` in the file named `hw06_predicted_probabilities1.csv`, and the predicted posterior probabilities of these 1000 data points for the positive class obtained using `Algorithm 2` in the file named `hw06_predicted_probabilities2.csv`.

3. Calculate possible classification thresholds using the predicted posterior probabilities. (20 points)

   ```
   print(thresholds1)
   [0.0109611  0.02522855 0.0290539  ... 0.964771   0.975904   0.9904605
   ```

   ```
   print(thresholds2)
   [0.00126964 0.00377441 0.00527235 ... 0.995633   0.9965655  0.9985055 ]
   ```

4. Calculate FP and TP rates using the true labels, predicted posterior probabilities, and thresholds. (40 points)

   ```
   print(fp_rates1[495:505])
   print(tp_rates1[495:505])
   [0.17611336 0.17611336 0.17611336 0.17408907 0.17206478 0.17004049
    0.17004049 0.17004049 0.17004049 0.16801619]
   [0.82608696 0.82411067 0.82213439 0.82213439 0.82213439 0.82213439
    0.8201581  0.81818182 0.81620553 0.81620553]
   ```

   ```
   print(fp_rates2[495:505])
   print(tp_rates2[495:505])
   [0.28744939 0.2854251  0.28340081 0.28137652 0.28137652 0.27935223
    0.27732794 0.27732794 0.27732794 0.27530364]
   [0.7173913  0.7173913  0.7173913  0.7173913  0.71541502 0.71541502
    0.71541502 0.71343874 0.71146245 0.71146245]
   ```
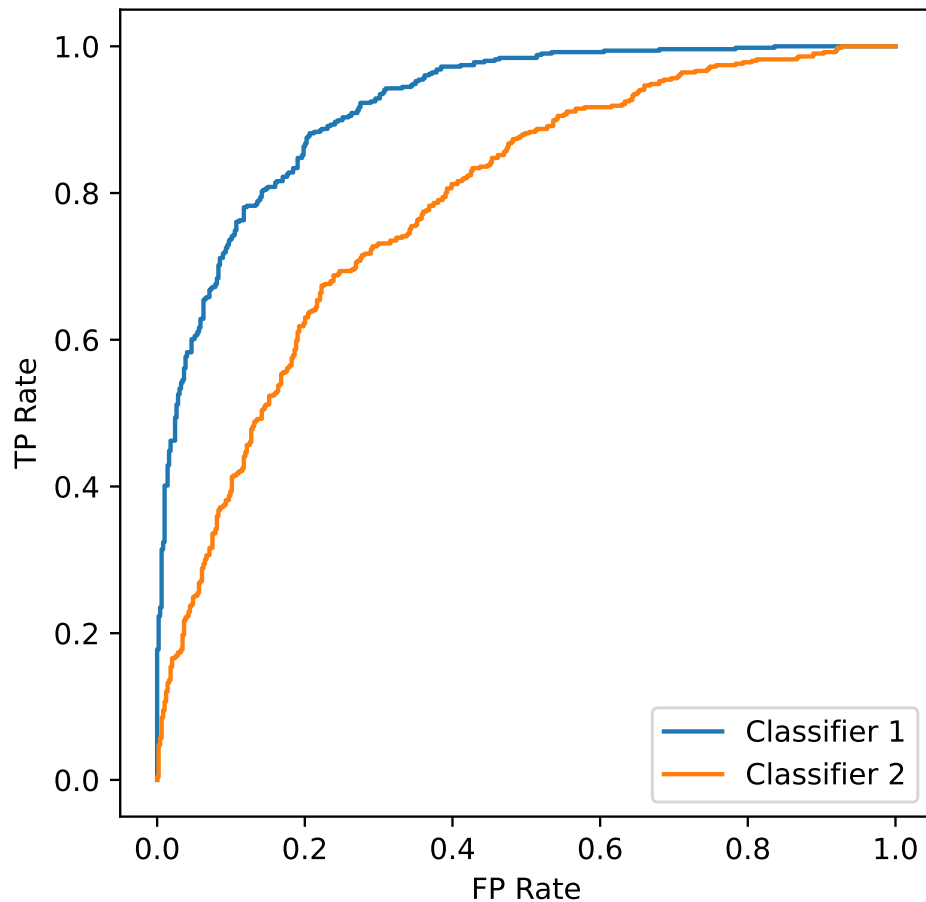
5. Calculate the area under the ROC curve using the FP and TP rates. (40 points)

```
print("The area under the ROC curve for Algorithm 1 is {}.".format(auroc1))
The area under the ROC curve for Algorithm 1 is 0.9183042358099563.
print("The area under the ROC curve for Algorithm 2 is {}.".format(auroc2))
The area under the ROC curve for Algorithm 2 is 0.7833928085644314.
```

What to submit: You need to submit your source code in a single file (`.py` file). You are provided with a template file named as `0099999.py`, where `99999` should be replaced with your 5-digit student number. You are allowed to change the template file between the following lines.

```
# your implementation starts below

# your implementation ends above
```

How to submit: Submit the file you edited to LearnHub by following the exact style mentioned. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.