

ENGR 421/DASC 521: Introduction to Machine Learning

Homework 6: ROC and PR Curves

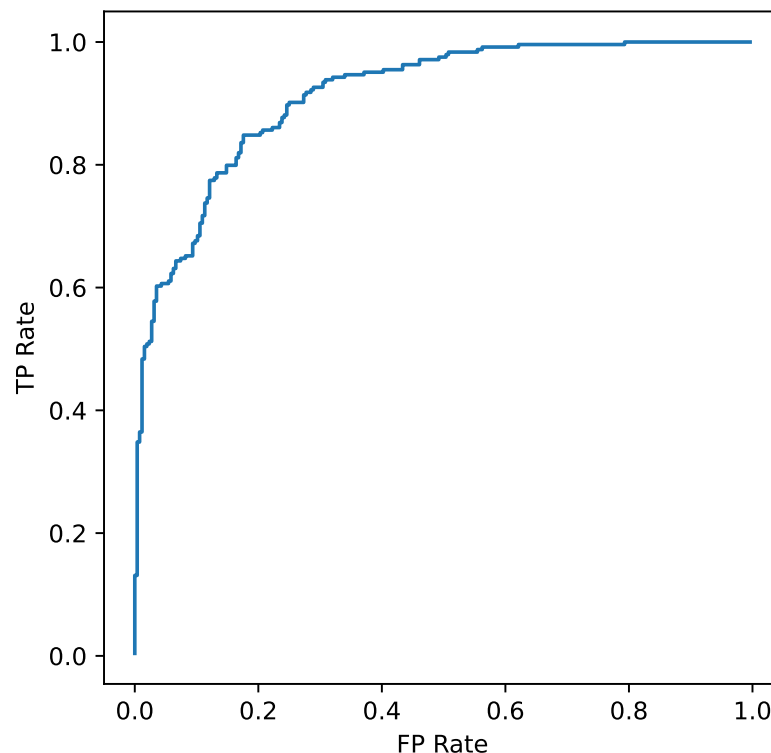
Deadline: January 20, 2023, 11:59 PM

In this homework, you will implement the receiver operating characteristics (ROC) and precision-recall (PR) curves 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 500 data points for a binary classification problem in the file named `hw06_true_labels.csv` and the predicted posterior probabilities of these 500 data points for the positive class in the file named `hw06_predicted_probabilities.csv`.
3. Implement a Python function with the following signature:

```
def draw_roc_curve(true_labels, predicted_probabilities):
```

where you draw the ROC curve and calculate the area under the curve. Your figure should be like the following figure, and you should report the area under the curve similarly. (50 points)

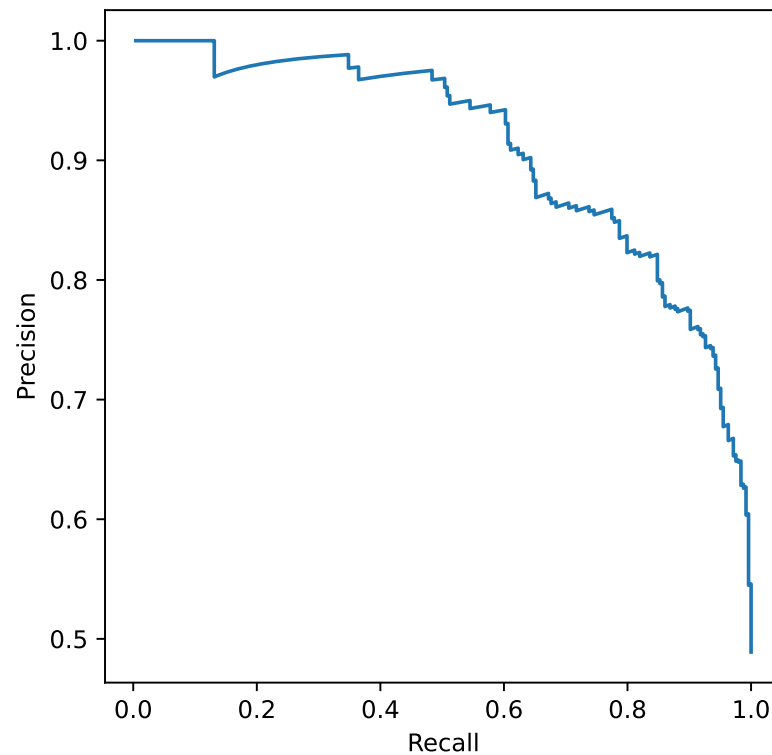


The area under the ROC curve is 0.9131499743852444.

4. Implement a Python function with the following signature:

```
def draw_pr_curve(true_labels, predicted_probabilities):
```

where you draw the PR curve and calculate the area under the curve. Your figure should be like the following figure, and you should report the area under the curve similarly. (50 points)



The area under the PR curve is 0.9104497751913916.

What to submit: You need to submit your source code in a single file (.py file) named as `STUDENTID.py`, where `STUDENTID` should be replaced with your 7-digit student number.

How to submit: Submit the file you created to Blackboard. Please follow the exact style mentioned and do not send a file named as `STUDENTID.py`. 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.
