IDSN 542: Machine Intelligence

Homework 5 Due: 10/22/25

Goal

Find a classification dataset. You can use Kaggle, UCI, or elsewhere.

Setup

- Your Python file(s) must begin with comments in the following format (replace the name and email with your actual information):

```
'''
Your name
IDSN 542, Fall 2025
Your USC email
Homework 5
'''
```

Requirements

Use the code provided from chapter 3 of the Geron book – as we covered week 8 in class. You can choose a binary classifier or a multiple-class classifier. It doesn't have to be images.

Compute the precision, recall, and f1 scores using the default SciKit-Learn values.

Generate the precision-recall curve.

Generate the ROC curve.

Normalize the confusion matrix and generate the plot of the errors (the last chunk of code from Thursday's lecture).

Write up a report as to what you did. This includes:

- Describe your dataset. Explain it in your report. This establishes context for me.
- What data preparation, if any, did you have to do.
- For the dataset you selected, is precision or recall more important – or is a balance of both? Why is that?
- What does the error plot indicate as to where you might want to augment/correct the data (like the last couple of slides from week 8 Thursday's lecture.

Submit your report and all your Python code in a ZIP file and submit to Brightspace.

This assignment is worth 30 points.