CSCI 3203: Tashfeen's Machine Learning I

Homework 2

You're only allowed to use the standard built-in Python 3 libraries with the exception of Numpy and Matplotlib.

Question 1. Plot the following trigonometric functions for $-5\pi \le x \le 5\pi$.

$$f(x) = x\sin^2(x), \quad g(x) = -x\sin^2(x)$$

- The title of the plot should read "Trigonometric Functions."
- The x–axis label of the plot should read "x–axis."
- The y–axis label of the plot should read "y–axis."
- The legends should appropriately denote the different functions by their definition.
- The graphs should have smooth continuos curves.

Question 2. Download and save the novel "The Cosmic Computer" by Piper. Read the saved text file into Python and plot a bar graph of the character frequency (case insensitive).

- The title of the plot should read 'Character Frequency in "The Cosmic Computer" by Piper.'
- The x–axis label of the plot should read "Alphabet."
- The y–axis label of the plot should read "Frequency."
- The bars should be sorted in ascending order.

Question 3. Using `matplotlib.pyplot.imshow` and an eight by eight binary Numpy array, plot the chess board pattern.

- The x–ticks should be letters from A to H, A being at the left.
- The y–ticks should be numbers from 1 to 8, one being at the bottom.
- The title should be "Chess Board Pattern."

Question 4. Following are the two matrices from the last homework,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 4 & 2 & 2 \end{bmatrix}$$

Use Numpy to multiply them and print the product $\mathbf{AB}$. Does it match with the one you carried out by hand?

Question 5. The Gregory series is given as,

$$f(n) = \sum_{i=1}^{i=n} \frac{4(-1)^{i+1}}{2i-1} = \frac{4}{1} - \frac{4}{3} + \frac{4}{5} - \cdots + \frac{4(-1)^{n+1}}{2n-1}$$

1) Without using any loops, i. e., using vectorised operations over Numpy arrays, plot the following values,

$$\{(n, f(n)) : 1 \le n \le 100\}$$

2) Plot the following error curve,

$$\{(n, (\pi - f(n))^2) : 1 \le n \le 100\}$$

Label the axis and title each plot appropriately. What is $\lim_{n \to \infty} f(n)$?

Question 6. Consider the following study results in table 1,

Point	Real	Predicted
1	−	+
2	−	−
3	−	−
4	−	+
5	+	−
6	+	+
7	+	+
8	+	+

TABLE 1. Real and Predicted Labels

State all the points that are,

1) True Negatives
2) True Positives
3) False Positives
4) False Negatives

Question 7. Give the confusion matrix for table 1.

	Predicted −	Predicted +
Real −	TN	FP
Real +	FN	TP

TABLE 2. Binary Confustion Matrix

Question 8. Again, for table 1, give the following in exact fractions.

1) Baseline Accuracy
2) Accuracy

$$\frac{\text{Total Correctly Classified}}{\text{Total Examples}}$$

3) Precision

$$\frac{\text{True Positives}}{\text{Total Examples Classified as Positive}}$$

4) Recall

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

5) F1 Score

$$\frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

Question 9. Often (but not always) how are precision and recall related?

Question 10. Between precision and recall, which is a better metric for,

1) Predicting good restaurants.
2) Predicting cancer.

Question 11. Between precision and recall, which metric is used in the Receiver Operator's Curve (ROC). What does the area under this curve tell us about a prospective classifier?

Submission Instructions

1) Submit a PDF that answers all the questions and includes any plots that the assignment asks for. Circle and/emphasize your final answer wherever possible.
2) Submit a Python file, e. g., `plot.py` that generates all the plots one by one.

Oklahoma City University, Petree College of Arts & Sciences, Computer Science