Homework 1: Mathematical Preliminaries
CPE 490/590 ST

Instructor: Rahul Bhadani

Due: January 22, 2025, 11:59 PM
100 points

You are allowed to use a generative model-based AI tool for your assignment. However, you must submit an accompanying reflection report detailing how you used the AI tool, the specific query you made, and how it improved your understanding of the subject. You are also required to submit screenshots of your conversation with any large language model (LLM) or equivalent conversational AI, clearly showing the prompts and your login avatar. Some conversational AIs provide a way to share a conversation link, and such a link is desirable for authenticity. Failure to do so may result in actions taken in compliance with the plagiarism policy.

Additionally, you must include your thoughts on how you would approach the assignment if such a tool were not available. Failure to provide a reflection report for every assignment where an AI tool is used may result in a penalty, and subsequent actions will be taken in line with the plagiarism policy.

Submission instruction:

Submission instruction for this homework supersedes one mentioned in the Syllabus.

This homework requires all answers recorded in a single `.ipynb` Python notebook. You can use a combination text cell (i.e. markdown formatted cell) and code cell to provide your answer. To add equations you should be able to use Latex syntaxs in the text cells of your Python notebook. As a part of your submission, you must provide executed notebook with code, text, and outputs. Alternatively, you can also provide a url (whose permission you must change to 'anyone with link can view') of your Python notebook from Google Colab. The naming convention for your notebook should follow the format {firstname.lastname}_CPE 490/590 ST_hw01.ipynb. For example, if your name is Sam Wells, your file name should be sam.wells_CPE 490/590 ST_hw01.ipynb.

Python Practice

Note: Python community has adopted following import convention that we will follow for the rest of the course:

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import statsmodels as sm
```

1   The Longest Substring (10 Points)

Write a function `LongestSubstring` in Python that takes a Python string as an argument. Your function should perform the following task:

1. If the string is empty print 'empty string' and return.

2. If the string length exceeds 20 characters, print 'maximum length exceeded' and return.

3. Otherwise, find the long substring in the given string that contains same characters. For example, if the input string is '2242777232888823', then the output should be '8888'.

4. Write code to use the function `LongestSubstring` as well.

The implementation doesn't need to be the most efficient implementation, and you may choose to use nested `for` loops as well.

2   Data and Time in Python (5 Points)

The built-in Python `datetime` module provides `datetime`, `date`, and `time` types. The `datetime` type combines the information stored in date and time.

Write a Python program that creates an empty file with the file name in the format ML-HW01_%Y-%m-%d-%H-%M.txt, where %Y is substitute for current year, %m is the substitute for current month, %d is substitute for current day, %H is the substitute for current hour, and %M is the substitute for current minute. One possibility can be a file name **ML-HW01-2011-10-29-20-30.txt**.

3   Working with Data Types in Python (5 Points)

In Python, there are some special data types such as table, and lists. A tuple is a fixed-length, immutable sequence of Python objects which, once assigned, cannot be changed. An example is `(2, 4, 5)`. On the other hand, lists are variable length and their contents can be modified in place. An example is `[1, 4, 5]`. Write a Python function that takes an argument as a list or a tuple, and returns a tuple containing the number of element the list or a tuple has, the maximum value, and a minimum value.

4   Slicing (5 Points)

You can select sections of most sequence types by using slice notation, which in its basic form consists of `start:stop` passed to the indexing operator `[]`:

In this part of the homework, you will create a 4×4 tensor using PyTorch package. Initialize your tensor randomly with only 0s and 1s. Tensor is a generalized name of a multi-dimensional array or matrix in machine learning domain. Note that this concept of tensor differs from one seen in Mathematics textbooks. Write a Python program to slice the last column of a 4×4 tensor.

5   Dictionary (5 Points)

Dictionary is a key-value pair data-types in Python. Create a dictionary in Python consisting of two letter state code of the neighboring states of Alabama as the key and their capital as the value.

Linear Algebra Practice

6   Linear Dependence (5 Points)

Let

$$A = \{(1,2,3),(2,3,4),(3,4,5)\} \tag{1}$$

$$B = \{(1,2,3), (2,3,1), (3,1,2)\} \tag{2}$$

Write the system of equations to test the linear independence of A and B and solve them using sympy package from Python.

7 Diagonal Matrix (5 points)

Write a function `ToDiagonal` in Python that accepts a square matrix (declared as numpy array of size $m \times m$) and returns an equivalent diagonal matrix. You also need to write code-snippet that uses the function `ToDiagonal`.

8 Solving a Matrix Equation (5 Points)

We have

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} \tag{3}$$

and

$$B = \begin{bmatrix} a & b \\ c & d \\ 2 & 1 \end{bmatrix} \tag{4}$$

such that

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5}$$

Use sympy Python package to write python code to compute the value of a, b, c and d.

9   Adjoint and Transpose (5 Points)

Using `numpy` Compute the transpose and adjoint of A where

$$A = \begin{bmatrix} 1 + 2j & 2 + 3j & 3 + 4j \\ 2 + 3j & 3 + 4j & 4 + 5j \end{bmatrix}. \tag{6}$$

Probability & Statistics Practice

10   Disjoint Sets (10 points)

If $P(A) = \frac{1}{3}$ and $P(B^c) = \frac{1}{4}$, can A and B be disjoint? Explain.

11   Coin Toss (10 points)

Write a Python program to simulate the tossing of a coin. After 100 tosses what is the probability of getting a head? What about after 1000 tosses? What about after 10000 tosses?

12   Three coins toss and CDF (10 points)

Consider the experiment of tossing three fair coins, and let $X =$ number of heads observed. The CDF of X is

$$F_X(x) = \begin{cases} 0, & -\infty < x < 0 \\ \frac{1}{8}, & 0 \leq x < 1 \\ \frac{1}{2}, & 1 \leq x < 2 \\ \frac{7}{8}, & 2 \leq x < 3 \\ 1, & 3 \leq x < \infty \end{cases} \tag{7}$$

Write a Python program to draw a graph $F_X(x)$ using step functions. Determine $F_X(2.5)$ using Python.

13 Probability Mass Functions: Poisson Distribution (10 points)

Refer to the Python Notebook provided at the end of Lecture 03.

Write a Python program to generate 10000 random samples that follow a Poisson distribution. You should choose values of λ as 0.25, 1.0, and 3.0. Plot the probability mass function for each the case overlaying on the same graph. Choose colors in the graph wisely so as to make the visualization understandable. Comment on how changing λ changes the shape of the probability mass function.

14 Probability Density Functions: Poisson Distribution (10 points)

Refer to the Python Notebook provided at the end of Lecture 03. Write a Python program to generate 10000 random samples following a Normal distribution. Keep $\mu = 0.5$ fixed and vary the standard deviation $\sigma = 0.3$, 5.0, and 10.0. Plot the probability density function using histogram and kernel density estimation (KDE) plot. Comment on your observation about changing the standard deviation and its effect on the shape of the probability density function.