DBS Associate Data Scientist Assessment: Section A

Question 1

Given the joint PDF:

$$f_{XY}(x,y) = \begin{cases} \lambda\mu e^{-(\lambda x+\mu y)}, & x \geq 0, y \geq 0 \\ 0, & otherwise \end{cases}$$

a) Find the PDF of $W = \frac{Y}{X}$

b) Find P(X<Y)

Question 2

Given three independent binary classifiers ($C_i, i = 1,2,3$) with uncorrelated error rates

e1 = 0.1, e2 = 0.3, e3 = 0.3

a) Use Majority voting to find the expected error rate of the ensemble. (Hint: This is not the average value of the errors)

b) What can be inferred if the assumption of independence is relaxed on the errors?

Question 3

Classify the test instances by hand using a Mixed Naïve Bayes model (Categorical + Gaussian).

ID	X1	X2	X3	X4	Y
Training					
1	S	42.5	N	F	N
2	S	39.2	H	T	N
3	-	33.6	H	F	Y
4	R	-	H	F	Y
5	R	22.8	N	F	Y
6	R	15.4	N	T	N
Testing					
7	O	25.0	N	T	?
8	S	36.4	-	F	?

Question 4

Write a function ***get_indices*** (Including type annotations) which takes a list of tuples and returns a 2D list of indices. Each sub-list corresponds to the indices of all rows pointing to the same person. Rows point to the same person if any of their column entries are the same. You will be graded on readability, error handling, completeness, and complexity (Please provide appropriate docstrings - explicitly stating space & time complexity).

Note: Your function should work for any number of columns/ people and be tested for other unseen edge cases.

```
>>> data = [('id1', addr1, 'pw1'), ('idx', 'addr1', 'pwx'), ('idz', 'addrz', 'pwz'),
('idy', 'addry', 'pwx')]

>>> get_indices(data)
>>> [[0, 1, 3], [2]]
```

Question 5

a) Write a NeuralNetworkClassifier class which inherits the BaseMLP abstract class (provided in base.py) and implements the ***fit*** and ***predict*** methods.
b) Utilize a simple test case to show that your classification model converges on a 2-dimensional non-linearly separable dataset.

*** You may only use NumPy for this question.

Note:

- Please make use of all the given hyper-parameters (You may include additional hyper-parameters in your child class).
- You will be graded on correctness, design and originality.
- Include all relevant documentation for your code (e.g. For each input argument)

Deliverables:

1) README detailing your project structure and any other relevant details.
2) Requirements file.
3) A src directory containing your code.
4) A single Jupyter Notebook displaying your answers to the above problems (Import your code from src)
5) A test folder for your test cases.