CS5103: Applied Machine Learning, Autumn 2025
Sample Questions for the Mid-Sem Exam

Q1. Are Python 'modules' and 'packages' synonymous? [2 marks]

Ans. No, they are not. Python modules are any organizational units of Python scripts that can be imported. On the other hand, packages are a specific type of modules that can contain other modules inside them. Each package has an attribute called '__path__' that a non-package module does not have.

Q2. What is the difference between a machine learning model and a machine learning algorithm? [2 marks]

Ans. A machine learning model is a mathematical structure, e.g., an equation in case of SVMs or a network in case of artificial neural networks. A machine learning algorithm is a computer algorithm that updates the values of the model parameters during training.

Q3. What is the following Python code snippet doing? Is the assigned value of 'test_size' advisable? [2+1 = 3 marks]

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.6, random_state=42
    )
```

Ans. The given code snippet is splitting a (samples 'X', labels 'y') pair into a training set (X_train, y_train) and a test set (X_test, y_test). The parameter 'test_size=0.6' implies that 60% of the samples will be randomly selected and moved to the test set. The remaining 40% of the samples will be kept in the training set. The parameter 'random_state=42' is setting the seed value to 42 for reproducible 'pseudorandom' number generation. It ensures that

the same samples are ‘randomly’ selected every time we run this code snippet. Therefore, the sample selection is not truly ‘random’, it is ‘pseudorandom’.

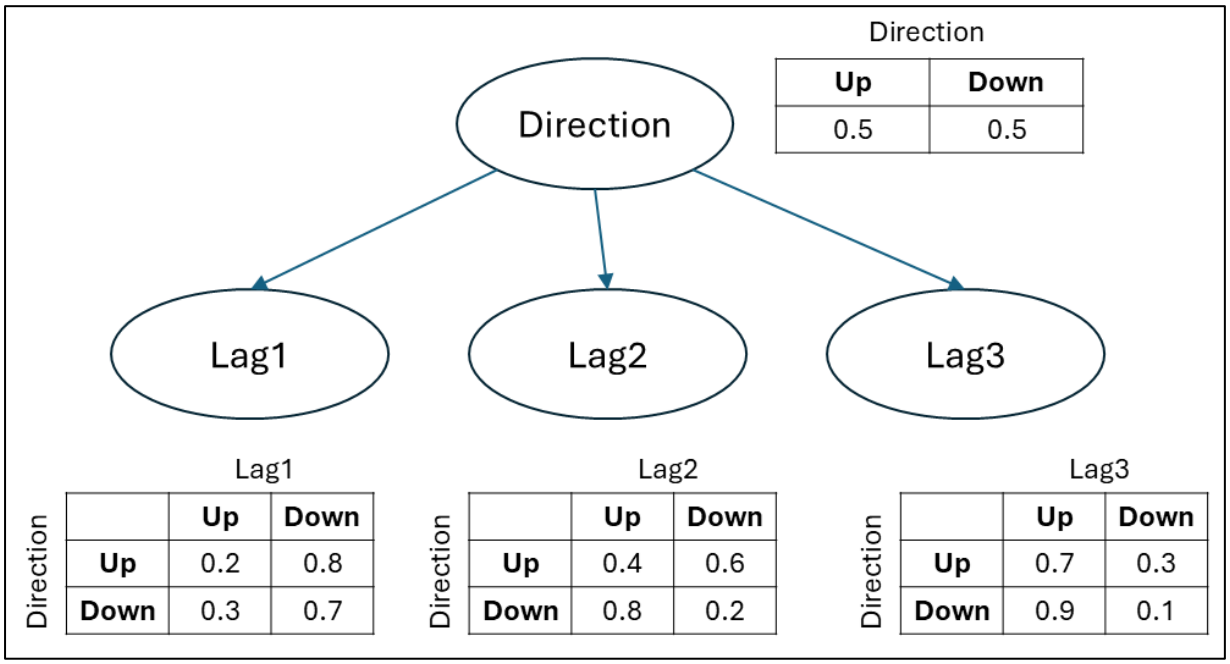
Moving 60% of the samples to the test set and keeping only 40% of the samples for training is usually not advisable. If we keep more samples (suppose, 80%) for training, the algorithm might be able to learn better and in turn perform better during test.

Q4. The following is a beautiful picture of the Aurora Borealis a.k.a. the Northern Lights, downloaded from https://www.pickpik.com/aurora-borealis-alaska-space-magical-night-aurora-australis-polar-lights-32294 . We would like to classify the Aurora Borealis by separating it out from its surroundings. Among the classifiers we have studied so far, please mention one classifier that could be useful to us for this purpose. [1 mark]

Ans. The Aurora Borealis looks like a parabola in this picture. Hence, we can use a quadratic classifier such as QDA. Otherwise, a support vector classifier with a polynomial kernel of degree 2 can also be used. In short, we need to apply a non-linear classifier here.

Q5. Suppose, we want to predict the 'Direction' of a share price on a particular day i.e. we want to predict whether the share price will go 'up' or go 'down' on that day. For the prediction, we would like to take the directions of the share price in the previous three days, represented by 'Lag1', 'Lag2', and 'Lag3', as inputs. Assuming we choose to use the naïve Bayes classifier, please draw the naïve Bayes model along with the model parameters (i.e. the conditional probability tables or CPTs). Initialise the CPTs with random values. [4 marks]

Ans.



Direction

Up	Down
0.5	0.5

Lag1

Direction	Up	Down
Up	0.2	0.8
Down	0.3	0.7

Lag2

Direction	Up	Down
Up	0.4	0.6
Down	0.8	0.2

Lag3

Direction	Up	Down
Up	0.7	0.3
Down	0.9	0.1

Exam Tip:

Divide the given time by the total marks to get an estimate of how much time you should spend for each mark. For example, if the question paper is of 40 marks and the given time is 90 minutes then there are only floor(90 / 40) = 2 minutes you can spend for each mark. This includes the time required for reading the question, writing the answer, and revising the answer.