ME 8813 Machine Learning Fundamental for ME
Homework Assignment 3
(Due 4/1/2022, Friday, 11:55pm)
(This is an individual assignment)

Electronic Submission Guideline: *your submission to Canvas* should be one python source file **ME8813ML_Homework3_YourLastName_FirstName.py** . or **ME8813ML_Homework3_YourLastName_FirstName.ipynb**, as well as a report file **ME8813ML_Homework3_YourLastName_FirstName_Report.pdf**.

Use the package Pomegranate to build a hidden Markov model (HMM). The documentation is available at: https://pomegranate.readthedocs.io/

The goal is to build an HMM to understand the development of high-tech companies based on their stock prices. The 10-year histories of stock prices for five companies are given. It is assumed that the life cycle for new technologies experiences five different states (**B**: Breakthrough, **A**: Adoption, **M**: Mature, **S**: Stagnant Economy, **D**: Decline), as shown in Figure 1.
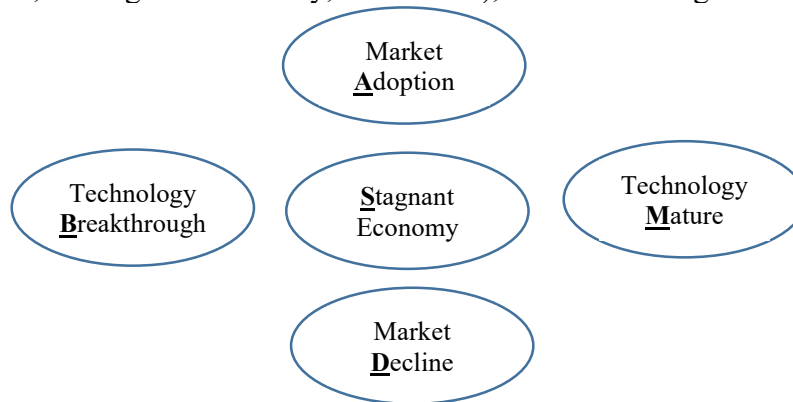

Figure 1. different states/stages of technology development

1) Based on the example demonstrated in class, construct an HMM with the above seven states and assign appropriate transition probabilities between the states. Assign each state an associated stock price change with a Gaussian/Normal distribution. For instance, we may have $N(0.01, 0.0001)$ with state "B", which indicates the stock price may increase 1% per day in average with variance of 0.0001 (standard deviation of 1%). For $N(-0.01, 0.0004)$ with state "D" or state "S", it indicates that the stock price may decrease 1% in average with variance of 0.0004. Given that the base prices for different companies are different, a normalized price with percentage is necessary.

2) Choose any three companies to train the HMM with the sequences of their stock prices with function `fit()`. The data include 10-year histories. You can choose a shorter period of time instead of 10 years to save computational time. The time period should be the same for different companies. The price to be used should be the close/last price of each day. ***Note that the training should be based on the percentage changes instead of the prices themselves directly.*** The training will update the initial probabilities that you assigned earlier.

3) Show the sequence of hidden states for the fourth and fifth companies based on their sequences of stock prices within the same time period. This can be done with function `predict()`.

4) Write a short report to show the graphical model of five states, the parameters (both transition and observation probabilities) after training, and the identified hidden state sequences for the fourth and fifth companies.