- Questions can involve a mix of writing code/scripts and answering questions or analyzing results.

- Code: Your scripts should be of the form q-x-y.py where x is the main question, y is the sub-question. For e.g., q-1-2.py is Python script for sub-question 2 within question 1.

- Ensure that submitted assignment is your original work. Please do not copy any part from any source including your friends, seniors and/or the internet. If any such attempt is caught then serious action will be taken.

- Use suitable train-validation split for your training and validation (20% of data).

- Numpy, pandas/csvReader(for data processing) are allowed.

- Report should contain details of algorithm implementation, results and observations.

1 Question

1. (70 points) Problem of Stock Prediction: You are given stock prices of Google Inc. of the past three years. The features you will be using are:
   1) Average of the low and high of the Google Inc. stock for the day.
   2) Volume of the stocks traded for the day.
   These will be used by you for predicting stock prices. Data set :
   `http://preon.iiit.ac.in/~sanjoy_chowdhury/GoogleStocks.csv`
   You are free to use library (Keras/ TensorFlow) for Part-1 and Part-2 of this question.

   1. **Part-1**: (30 points) **RNN**
      Perform stock prediction using RNN for the following configurations

Number of Hidden Layers	Number of cells in hidden layers	Number of Time steps
2	30,50,80	20,50,75
3	30,50,80	20,50,75

      Note: Get results for total 18 combinations from above table. For example, 1 combination will contain 2 hidden layers , the number of RNN cells will be 30 units

and the number of time steps will be 20. So this combination will be denoted as RNN (2,30,20).

2. **Part-2**: (30 points)**HMM**
Perform stock prediction using HMM where you will need to vary

Number of Hidden States	Number of Time steps
4,8,12	20,50,75

Note: Get results for total 9 combinations from above table. For example, 1 combination will contain 4 hidden states, and the number of time steps will be 20. So this combination will be denoted as HMM(4,20).

3. **Part-3** (10 points) Compare the results of any one combination from RNN with any one combination from HMM. All results will be time series plots which will also have the plot of the original data.

2   Question

2. (30 points) Given a DNA sequence and a state path sequence, find the probability that the given DNA sequence is generated from the given path sequence only.

[Image source: https://biointelligence.wordpress.com/tag/hidden-markov-models/]
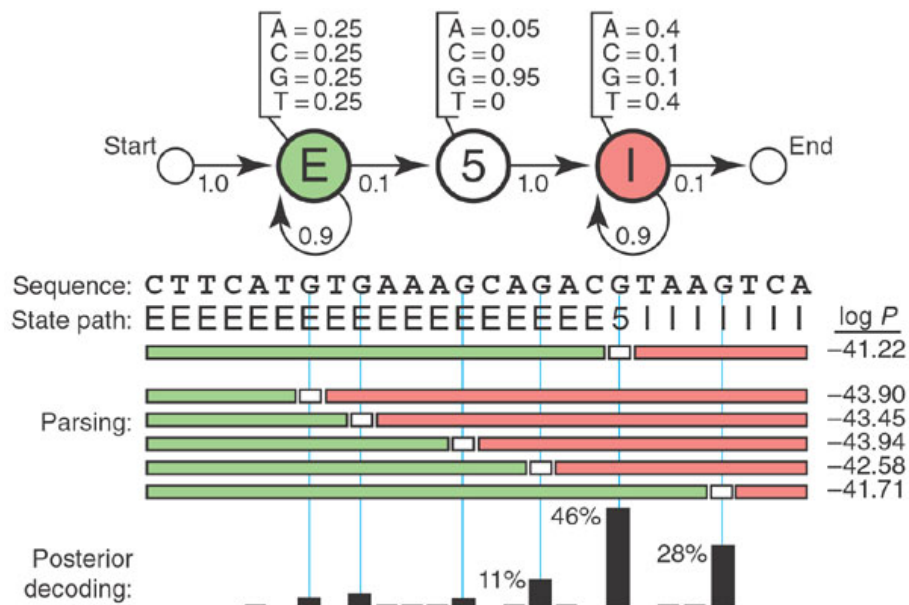
- Input DNA sequence: "CTTCATGTGAAAGCAGACGTAAGTCA"
- Input State path sequence: "EEEEEEEEEEEEEEEEEEEE5IIIIIII$"

- Emission & Transition probabilities can be found from above figure

- Output: State path probability for above state path [Log probability required i.e. log(p)]