MA691 : Endsem Exam

Instructor: Dr. Arabin Kumar Dey

1 Notes:

- Please do not violate the honor code.
- Use Python or R to make your code.
- Upload a separate report file (handwritten or typed out) to answer all the questions in text. Please submit all documents (Code, report etc) in Google form sent in your mailbox.
- Total Marks : 30; Exam Duration : 2 hr.

2 Questions:

1. Write python/R code to drag a dataset of size 100 for Bitcoin prices using Quandl (or any platform). Implement usual LSTM model to make 20 day forecast. Sketch the computational graphical structure for a highway type LSTM model for implementing back propagation. Use your class note on derivation of backpropagation on usual LSTM to extend back propagation in highway type LSTM. Provide some implementation approach for highway LSTM both in latest version of Tensorflow and Torch Environment.

   $[1 + 2 + 2 + 4 + 5 = 14]$

2. Write python/R code to drag the dataset of 100K usermovie dataset. Filter the sparse data by ordinary averages of the rating provided by each users. Use a RBM and DBN machine to extract the user features. Use KNN or Kmeans algorithms to cluster the datasets using the feature vectors obtained through DBN and RBM. [The process can help you to figure out the number of hidden user or movie features] [2 + 2 + 4 + 2 = 10]

3. Local trend models can be specified as ; $y_t = \mu_t + \epsilon_t$, $\mu_{t+1} = \mu_t + \eta_t$ where, $\epsilon_t \sim N(0, 5)$ for each t and $\eta_t \sim N(0, 2)$ for each t. ϵ_t and η_ts are independent.

   Assume that y_t's are logarithms of intradaily realized volatility of Alcoa stock from January 2, 2003 to May 7, 2004. The realized volatility is computed from the intraday 10-minute log returns measured in percentage.

   (a) Plot the actual realized volatilities using the above dataset obtained from the TAQ database of the NYSE. Use both Kalman Filter and particle filter to extract the volatilities based direct python module or using your code.

   [4 marks]

4. Consider the toy example : Three Urns (U_1, U_2 and U_3) are filled with different proportion of three colors of balls (Red (R), Green(G) and Blue(B)). Let's assume you are using the following set of scheme to obtain a sequence of colors { R, R, G, G, B } : Select an Urn with probability distribution {0.6, 0.3, 0.1}. Pick up a ball at random from the Urn. Note down the color of the ball and return it back to the urn. With a switching probabilty choose the next urn and again pick a ball at random from the chosen urn and so on.

   Let's assume that Switching probability is determined by the following matrix :

$$\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

and number of balls available in three urns are as follows : Urn 1 : 70 G, 20 R, 10 B; Urn 2 : 50 G, 20 R, 30 B; Urn 3: 40 G, 40 R, 20 B.

Use suitable python code module to extract the hidden sequence of Urns.

[2 marks]