CS223 Module 2 Week 8 Programming Assignment

This week's programming assignment actually consists of two programming assignments. They each count 40pts. You will need all files in the Module 2 Week 8 Code folder on Canvas. Unzip the file named pydotplus.zip and place the folder in your current Python working directory. Or you can do a `pip install pydotplus` to install the package.

Module 2 Week 8 Programming Assignment #1:

This assignment involves modifying the hmm code in the file named hmm.py. The example code you will be running and extending for this programming assignment is taken from Jason Eisner's Ice Cream and Baltimore Summer example that can be found at http://www.cs.jhu.edu/~jason/papers/#eisner-2002-tnlp . The current hmm model in the hmm.py code accounts for just cold and hot days. Eisner eats probably just 1 cone on cold days and likely 3 cones on hot days, but possibly eats 2 cones on cold or hot days.

The modification you need to make is to add a third typr of day, called warm so that now you have cold, warm, and hot days. Now Eisner will likely eat 1 cone on cold days, 2 cones on warm days, and 3 cones on hot days. Adjust the transition and emission probabilities as you desire. Document your code to clealy indicate these probabilities so that it is easy for the instructor to find.

The run the modified hmm code to predict the underlying mm for the same training and test sequences as used in the original code.

Submit your .py code to Programming Assignment #2 in Canvas.

Module 2 Week 8 Programming Assignment #2:

This assignment involves using, at your discretion, either the code in the hmm.py file or code in the hmm_example.py file to implement a hmm that will find CpG islands in the DNA sequence that is specified in the file named dna_seq.txt that is located iin the Module 2 Week 8 Data Sets folder.

Use the course HMM slide decks, YouTube video presentations, and provided publications to design and build your HMM CpG island detecting model. Use the letter "I" for island and "N" for non-island. Your output shod be as follows
A list of lists where the first list is the original DNA sequence provided in the dna_seq.txt file, and the second list contains "I"s and "N"s indicating where your HMM model predicts whether the corresponding nucleotide is part of ("I") or not part of ("N") a CpG island. For example
[gcgtataggcgattcggat …]
[IIINNIIINNNNIIINIII …]

Submit your .py code to Programming Assignment #3 in Canvas