CS582 Final
Due Friday, 05/15/15, 12:00am

Notes:

1. Form as well as content will be graded, so express your answers correctly, clearly and concisely utilizing complete sentences, thoughts and ideas.
2. Please do not submit scanned material as it is difficult to read and cannot be commented.
3. Submit your answers as an MS-Word document by email to me at: CKonopka@mail.sdsu.edu
4. This test is worth 25% of your total grade for the semester.
5. The extra credit is worth ***up to*** 10 *additional* points.
6. This test is due on **Friday 05/15/14** at 12:00 am (midnight) – not later!

Advice:

Claiming the work of another as your own is plagiarism and so we (of course) attribute referenced works to avoid this issue. However, the purpose of taking a test should be more than merely following the rules – and must more than mere recitation. Consider this exam as an opportunity for you to express your mastery of the topics presented in this class. Short, high-level answers aren't worth nearly as much as answers which demonstrate your profound understanding of a given topic. Longer written responses will of course take more time to compose - and you'll have plenty of time. The best strategy is simply not to delay. Finish an early draft now, if possible, then take a break before reviewing, correcting and improving prior to submission. Please don't hesitate to email me with any questions. That's why I'm here and it's why I frequently check my email. Also, don't be afraid to collaborate with your fellow students. It's OK to discuss exam topics, as long as your work remains your own and not a copy or a slight re-phrasing of another's. Finally, good luck!

Questions:

1. **(10 Points)** Feature Extraction
    a. (5 pts) Describe the Fourier Transform. Be sure to include:
        i. What happens when the Fourier transform's data assumptions are violated?
        ii. What is overlap and why is it used?
        iii. What is the time-frequency resolution trade-off and how is it addressed?
    b. (5 pts) Calculate the frequency bin width of a 256-point Fourier vector where the data is sampled at 22,000 Hz.
2. **(10 Points)** Describe the Wavelet Transform. Be sure to include:
    a. (2.5 pts) A description of the continuous wavelet transform;

b. (2.5 pts) A description of the discrete wavelet transform;

c. (2.5 pts) A description of the wavelet packet transform;

d. (2.5 pts) Compare the Wavelet and Fourier Transforms;

a. **(20 Points**) Describe in detail the three issues related to the use of the Hidden Markov Models:

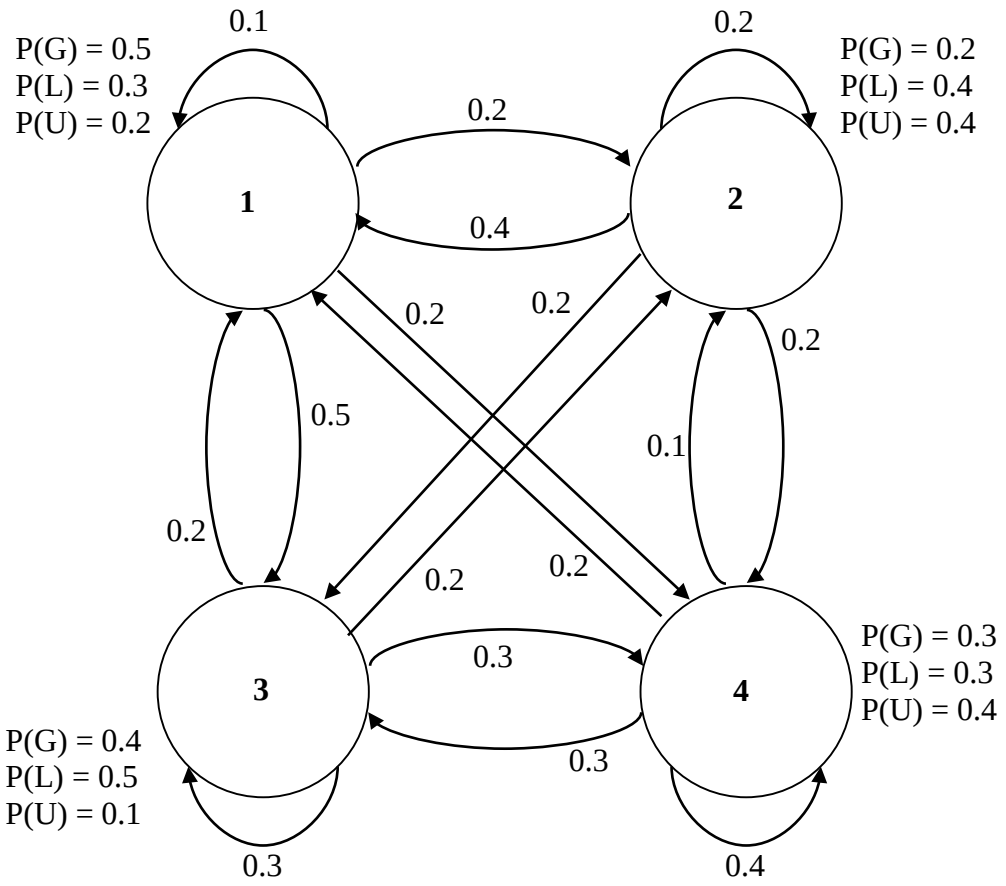
e. (5 pts) Evaluation;

f. (5 pts) Traversal;

g. (10 pts) Training;

3. **(20 Points) Note:** for this problem you may provide either a hand-calculation or use the Matlab code you developed for the HMM assignment. Either way, you must show your work – i.e. the calculations if done manually, or a sample run and the code if done using your code. Given a Hidden Markov Model trained to recognize certain stock market trends where G = a market gain, L = a market loss and U = an unchanged market.
   a. (10 pts) Calculate $P(O|\lambda)$ for O = (U,U,L,L,U,U,U,G,G,U,U,LL,U,U,U,G,G) and $\lambda = <A, B, \Pi>$,
   b. (10 pts) Calculate the Viterbi Score and best path for the same observation sequence (O)

   Where:



State Transition Probabilities

A	1	2	3	4
1	0.1	0.2	0.5	0.2
2	0.4	0.2	0.2	0.2
3	0.2	0.2	0.3	0.3
4	0.2	0.1	0.3	0.4

Output Probabilities

B	1	2	3	4
P(G)	0.5	0.2	0.4	0.3
P(L)	0.3	0.4	0.5	0.3
P(U)	0.2	0.4	01	0.4

Initial State Probabilities

π	1	2	3	4
π_i	0.25	0.25	0.25	0.25

4. **(20 Points)** Describe the following Hidden Markov Models:
    a. (5 pts) Discrete;
    b. (5 pts) Semi-Continuous;
    c. (5 pts) Continuous;
    d. (5 pts) Explain how Hidden Markov Models process real-valued feature data (e.g. MFCC feature vectors).
5. **(20 Points)** Grammar: Define and discuss the following:
    a. (4 pts) Noam Chomsky's theory of "Universal Grammar"
    b. (4 pts) N-gram modeling;
    c. (4 pts) Smoothing, why and how it is used;
    d. (4 pts) Language perplexity;
    e. (4 pts) Semantics: meaning, representation, implementation, etc;

Extra Credit:

Note: The following is worth ***up to*** 10 **points.**

1. Sentence generation:
    a. Select one or more “text corpora” the Project Gutenberg site is a good source, but any source of text material will due.
    b. Generate a n-gram model using this data source;
    c. Generate sentences using this model;
    d. **Notes:**
        i. You may combine the work of related authors, examples might include:
            1. song lyrics;
            2. Poetry from similar poets;
            3. Fictional works written in a similar style, era, etc. (e.g. “String Theory” papers, Beat Poets, Victorian Era writers.
    e. Turn in:
        i. A detailed discussion of your work;
        ii. A text file containing the concordance;
        iii. Sample runs;
        iv. The code and any other relevant files;