SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

01.112 Machine Learning, Spring 2018
DEEP Exam

Date/Time: 27 April 2018 19:00 – 28 April 2018 19:00

Instructions:

1. This is a closed book exam.
2. You are allowed to use non-programmable calculators.
3. You may NOT refer to any material.
4. You may NOT access the Internet.
5. You may NOT communicate via any means with anyone.
6. Please submit your solutions to eDimension before the deadline. No late submissions would be accepted.

Question 1. (*25 points*)

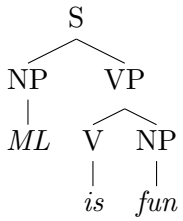
In class we discussed softmax regression and conditional random fields. We mentioned that logistic regression can be regarded as a special case of softmax regression or conditional random fields. Formally show why this argument is correct.

Question 2. (*25 points*)

In class we discussed structural SVM for structured prediction. Actually structural SVM can be regarded as an extension to the classic binary SVM that we discussed in class. Discuss how to use structural SVM to perform multi-class classification.

Question 3. (*25 points*)

In structured prediction, mapping a linear structure to a tree structure is a more advanced class of problem called probabilistic parsing. A classic model here is a generative model called probabilisitic context free grammar (PCFG). Turns out there are elegant algorithms for doing parsing efficiently for both decoding and inference (the corresponding algorithms are analogous to the Viterbi and forward-backward algorithms for HMM). For example we are given a sentence *ML is fun*, we would like to parse the sentence into the following tree structure for a given probabilistic grammar below:



- S → NP VP (0.1)
- S → NP PP (0.8)
- S → NP (0.1)

- NP $\rightarrow$ *ML* (0.1)
- NP $\rightarrow$ *fun* (0.2)
- NP $\rightarrow$ *job* (0.7)
- VP $\rightarrow$ V NP (0.7)
- VP $\rightarrow$ V (0.2)
- VP $\rightarrow$ V PP (0.1)
- PP $\rightarrow$ *is* (0.1)
- PP $\rightarrow$ *fun* (0.1)
- PP $\rightarrow$ *at* (0.8)
- V $\rightarrow$ *at* (0.2)
- V $\rightarrow$ *is* (0.3)
- V $\rightarrow$ *fun* (0.5)

Each grammar rule, analogous to the transition terms or emission terms, are responsible for generating a list of items to the right of the $\rightarrow$ with a probability. For example, the probability of generating the list "NP VP" from the symbol "S" is 0.1.

Clearly describe the algorithm for finding the most probable parse tree for the sentence "*ML is fun*" based on the above grammar.

Question 4. (*25 points*)

The probabilistic latent semantic analysis (PLSA) model is a generative model. One version of the model assumes the words in a document with n words $(x_1, \ldots, x_n)$ and the document label (y) are generated jointly based on the following formula:

$$p(x_1, x_2, \ldots, x_n, y) = p(y) \sum_{j=1}^{m} p(c_j|y) \prod_{k=1}^{n} p(x_k|c_j)$$

where c's are the latent variables, which are referred to as "concepts" or "topics". Describe an algorithm for learning the PLSA model.