LING572 Hw9: Neural Network
Due: 11pm on March 15, 2017

For this assignment, you will watch the 2013 NAACL tutorial (Part 1 and 2) and answer the following questions.

- The videos and slides are available at http://nlp.stanford.edu/courses/NAACL2013/.

- The videos are also available at https://www.youtube.com/watch?v=eixGKz0Asr8 (Part 1) and https://www.youtube.com/watch?v=zHXOHqIyeD4 (Part 2).

- All the slide numbers below refer to slides at http://nlp.stanford.edu/courses/NAACL2013/NAACL2013-Socher-Manning-DeepLearning.pdf

- There are 15 questions and each is 10 points. So the total raw score is 150 points. Your final grade for hw9 will be the minimum of 100 and the raw score. In other words, one can get the maximal score, 100 points, by correctly answering two thirds of the questions. Therefore, you can skip a question or part of a question.

- When watching the videos, it might be helpful to pause and replay to ensure that you understand the material before moving on. The videos are about 3 hours, but it might take you much longer to finish it.

- The videos will not cover all the slides. The questions below will not refer to slides that are not covered in the videos.

- The tutorial assumes that the audience are familiar with certain concepts such as sigmoid functions. Question Q1 and Q3 are about those concepts. The answers are not given in the lecture, but can be easily answered by googling those concepts. I include some urls for your convenience. You do not need to understand everything on those pages.

- Other than Q1 and Q3, the answers to other questions are all provided on the slides or verbally by the lecturers.

- The order of the questions may or may not follow the order of the lecture. Some questions refer to page numbers on the slides.

- Your answers to the questions should be concise and no more than a few lines. No need to go into too much detail.

Q1 (10 points): Logistic function and sigmoid function:

1a (4 pts): What is a logistic function? Write down the formula.
    See https://en.wikipedia.org/wiki/Logistic_function

1b (3 pts): What is a sigmoid function? Write down the formula.
    See https://en.wikipedia.org/wiki/Sigmoid_function

1c (3 pts): What is the relation between these two functions?

Q2 (10 points): Tanh function:

2a (5 pts): Write down the formula for the tanh function.

2b (5 pts): What is the relation between sigmoid function and tanh function?

Q3 (10 points): The softmax function:

3a (5 pts): What is the softmax function? Write down the formula.
See https://en.wikipedia.org/wiki/Softmax_function

3b (5 pts): If a vector x is [1, 2, 3], what is the value of softmax(x)?

Q4 (10 points): What is representation learning? Why is that important?

Q5 (10 points): What is distributed representation of words? Why does it work better than word clusters?

Q6 (10 points): Neural network was studied decades ago without achieving good results. Why is it more successful now?

Q7 (10 points): Matrix notation on Slide #29:

7a (3 points) : What is $\mathbf{W}$? Is it a number, a vector or a matrix? What are the dimensions of W, based on the figure on the slide?

7b (7 points): Same question as in 7a for $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{z}$ on that slide.

Q8 (10 points): What is one-hot representation of words? Is it a good representation of words? Why or why not?

Q9 (10 points): On slide #49, suppose we represent each word as a n-dimensional vector, the window size is k (e.g., the window size in that example is 5), and the hidden layer is h-dimensional. What are the dimensions of $\mathbf{x}$, $\mathbf{W}$, $\mathbf{a}$, and $\mathbf{U}$?

Q10 (10 points): What does unsupervised pre-training mean? Does it help system performance? Why or why not?

Q11 (10 points): On Slide #71, the training has two stages.

11a (4 points): What are the two stages?

11b (3 points): What is the purpose of the first stage? In other words, what is the output of that stage?

11c (3 points): Does Stage 2 change the word vector representation? And does changing the word vector representation improve or hurt system performance?

Q12 (10 points): RNN for parsing (Slide #100-111):

12a (3 pts): How is the meaning of phrases represented?

12b (7 pts): How can we use recursive neural network (RNN) to build a parser?

Q13 (10 points): Syntactically-untied RNN (SU-RNN) and CVG:

13a (5 pts): What is the difference between RNN and SU-RNN? What benefits do SU-RNN provide?

13b (5 pts): What is CVG? What benefits does it provide over using SU-RNN only?

Q14 (10 points): Common sense reasoning with knowledge base (see #173-176):

14a (3 pts): How are entities in the knowledge base represented?

14b (3 pts): How is a relation represented?

14c (4 pts): To train such a model, where do negative examples come from?

Q15 (10 points): Compared to the learning algorithms covered in LING572, what are the main advantages of neural network? Please list two or three.

Submission: Submit a tar file via CollectIt. The tar file should include the following.

- In your note file hw9-notes.*, include your answers to the questions, and any notes that you want the TA to read.