CST 463 - Advanced Machine Learning

Dr. Glenn Bruns

# Lab: Attention

**Please work with your team**. Don't look at the hint for a problem until your team agrees on your answer and the reasoning behind it.

1. (Fill in the blank) The context value used by the decoder is a weighted sum of \_\_\_\_\_\_\_\_\_\_.
2. (T/F) Attention in a Seq2Seq network could only be used if the encoder is bidirectional.
3. (T/F) In the Seq2Seq model in the lecture, both the encoder and decoder were bidirectional.
4. (T/F) The weights (symbol ) used to compute the context vector change at every step of the decoder.
5. In the slide 'Basic concept of attention', what kinds of values are used to represent the words in the source sentence?
   1. strings
   2. index values
   3. embeddings
6. (T/F) The weights used to compute a context vector are always positive.
7. The weights at a decoder step are computed from which values? You don't need to give symbols, just describe the values.
8. Intuitively, when is the weight value large?

### 

### Hints:

1. encoder output values
2. there's nothing that requires that encoder output values come from a bidirectional encoder
3. only the encoder is bidirectional in the model
4. Yes, the weights must change at each step.
5. Embeddings are used to represent the words.
6. The weights are softmax'ed, so they are always positive and sum to 1.
7. The weights are computed using the decoder input state at that time step, plus the output values at each encoder step.
8. The value is high if the decoder state vector at step i is similar to the jth encoder output value.