MSIA_490-0_SEC20 Lab 3 Discussion Questions

Skip-Gram

Q1: How is a Skip-gram model trained? What is its objective and how is softmax useful?

*Q2: What is the problem with the original Skip-gram model? How is it solved? (Conceptual discussion is enough.)

Q3: What problem is "Subsampling" trying to solve?

Bert

Q4: What makes Bert different from Skip-gram? (What tasks are they trying to solve?)

Q5: What are the advantages of pre-training?

*Q6: Which fine-tuning tasks may cause Bert to seem less robust?

Bonus

Q7: Would you call Elmo a language model? Why or why not?

*Q8: What is the advantage of the architecture of Bert comparing to the one of Elmo?