Today’s Schedule

• Frequently Asked Questions
• Taking more Questions
- We will not be testing degenerate cases. That is, you don’t have to worry about testing “bad” inputs.
- Make sure your types match exactly that of what is specified in the handouts and Piazza.
- Make sure that your functions can take in what is specified.
- Make sure your code works on CDF.
Q: Do we have to cover X?

A: If it is not specified in the handout/piazza answer's you do not have to cover it.
Task 1

Q: SENTSTART and SENTEND have to be all-caps?

A: Yes
Q: Are SENTSTART and SENTEND added in this stage?

A: Yes
Q: Perplexity is increasing?

A: What does delta > 0 mean? What happens if delta = 0? Look at the code to see how it handles certain cases. This will help explain.
Task 3

Note: Indentation is acting up on different OS’s for different people.

Fix the docstring by indenting 1 right (ONLY IFF DOESN’T COMPILE)

Fix the if statement at the very end if it isn’t outside the for loop. @467
Task 4

Q: Making the alignment matrix is too computationally expensive!

A: It is not meant to be a matrix. Rather a dictionary. If implemented correctly, it will be feasible.
Q: How long is 1k supposed to take?

A: ~Under a minute, and if your implementation is really slow then a couple of minutes, but definitely not very long.
Q. Step 4 initialization
If we have Sentences
SENTSTART A B SENTEND
SENTSTART X Y SENTEND

There are 3 potential options when we initialize AM
1. AM[“A”][“SENTEND”] = 0
2. AM[“A”][“SENTEND”] = 0.25
3. AM[“A”][“SENTEND”] entry does not exist
Intuition:
The only field for ‘SENTSTART’ should be ‘SENTSTART’
(and similarly for ‘SENTEND’).

Intuitively, the sentence start and end tags are independent of the languages, there’s no uncertainty here - we know 100% that the source start/end tag should be aligned with the target start/end tag.

For those who have already implemented your solutions with SENTSTART, SENTEND included in your alignment computation, we will accept the solution 2.
Q: Can any other word link to SENTEND? eg., AM[“the”][“SENTEND”] exits? or Should SENTEND and SENTSTART be removed for the AM and then added afterwards?

A: See previous answer. Should have been removed but we’ll accept non-removal.
Q. Why are the initialization methods different in the lecture slides and in the tutorial 2 handout?

A: Kindly read the ‘Note’ section of Task 4 in the handout. There’s 1. Naive Initialization’ and 2. Targeted Initialization.

1. Naive Initialization = \( \frac{1}{||V_f||} \)
2. Targeted Initialization = \( \frac{1}{||\text{number of unique (e,f) associations}||} \)

The lecture handout uses [1], the tutorial handout shows [2] - ergo they are different.

@450
Q: If we have the following pairs: (“the cat”, “la chat”) (“the dog”, “la chien”) what would be the AM for “the”?

AM[the][la] = 1/3
AM[the][chat] = 1/3
AM[the][chien] = 1/3

OR

AM[the][la] = 2/4
AM[the][chat] = 1/4
AM[the][chien] = 1/4

A: It’s the first one. Essentially 1 / set(targeted associations (e,f) )
Task 5

Q: Where is BLEU_score.py
A: @366
Q: Is the BLEU_score.py actually calculating the BLEU score?

A: No. It only calculates parts of it (depending on the parameters). For the complete BLEU score you will have to combine multiple calls to the file + add Brevity.
Q: Do we implement capping in BLEU score?
A: No.
Today’s Schedule

• Frequently Asked Questions

• Taking more Questions