CS 753: Assignment #3

Preethi Jyothi
pjyothi@cse.iitb.ac.in

IIT Bombay — October 28, 2019

Instructions: This is a relatively short assignment due on or before 11.55 pm on November 4, 2019. No grace period will be granted for this assignment. The submission portal on Moodle will be closed after 11.55 pm on November 4.

- For this assignment, you can work in groups of two or three.
- You should submit a pdf file named `submission.pdf` to Moodle, which is **no longer than 2 pages (11 pt font)**.
- Please make sure you understand this material well. One question based on RNN transducers will appear in the final exam.

Another end-to-end speech recognition model

Among end-to-end paradigms for ASR, we have learned about Connectionist Temporal Classification (CTC) based training and encoder-decoder models with attention. A third paradigm that is gaining traction in recent years is the *RNN Transducer* model [1]. This technique helps get around the frame-independence assumptions prevalent in CTC training and natually enables online streaming, which is difficult with the attention-based encoder-decoder model.

Question 1

(a) Clearly describe the RNN transducer (RNN-T) model. You can use a diagram to accompany your description. [**10 points**]

(b) How does RNN-T relax CTC's frame independence assumption? [**4 points**]

(c) How does RNN-T make it easier to do online streaming? (That is, it can continuously process input samples and stream output symbols.) [**4 points**]

(d) List one limitation of the RNN transducer model. [**2 points**]

References:

[1] A. Graves, "Sequence Transduction with Recurrent Neural Networks", https://arxiv.org/pdf/1211.3711.pdf, 2012.