Machine Translation Sabrina Brändle, Eirini Valkana

**MT Exercise 4  
Layer Normalization for Transformer Models**

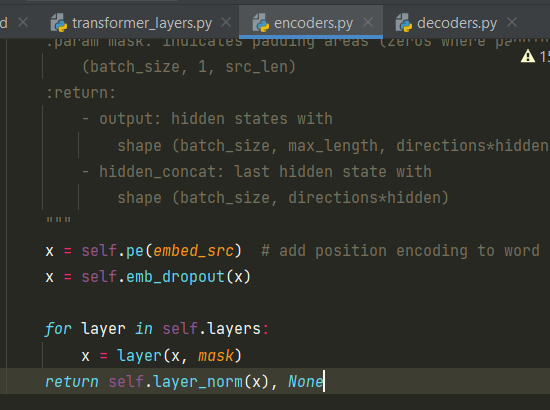
**Link to ‘ex4’ repository:** <https://github.com/sabrinabraendle/mt-exercise-4>

**Instances of LayerNorm in JoeyNMT**

* For each instance of layer normalization in JoeyNMT that you can find, describe whether it corresponds more to pre- or post-norm and which sublayer it pre- or post-modifies.
* Important: Make sure to clearly specify the python script and relevant code for each instance.

1. Python script: encoders.py

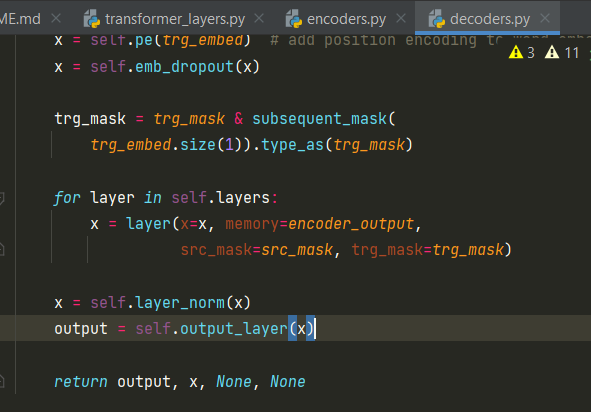
* relevant code:



* corresponds to: post-norm

1. Python script: decoders.py

* relevant code:

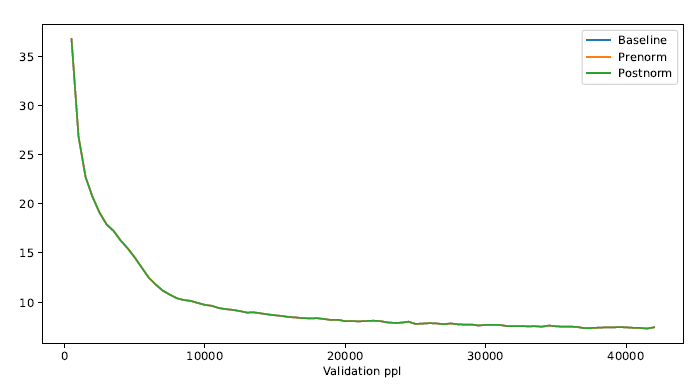


* corresponds to: pre-norm

**Your implementations**

* Modify files for pre- and post-norm – add & commit correct branches
* Check branch: git status / go to branch: git checkout [branch-name]
* Change model name / log file name (before training)
* (Hint: To make sure you are really using the correct version, use the logging functionality with which JoeyNMT prints out logging information. You can for example add a print statement at an appropriate place)
* Run ‘./scripts/train.sh’ for each model

**Discussion**



Discuss your results using the line chart. In your discussion, make sure to comment on the following points:

* Given that there is a difference in the training progress for the three models, can you think of a reason for it?
* In what way does our setup differ from Wang et. al. 2019? How could that have influenced our results?