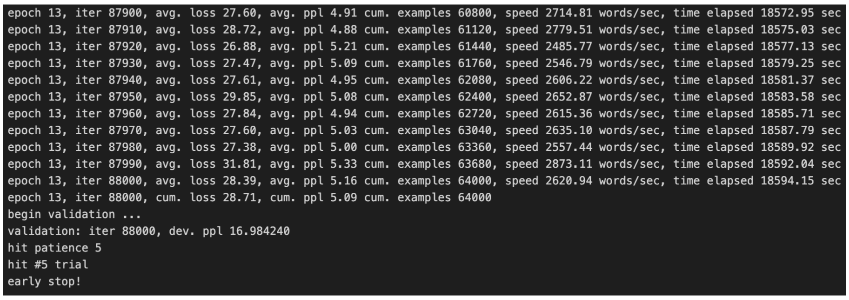
cs224n Assignment #4

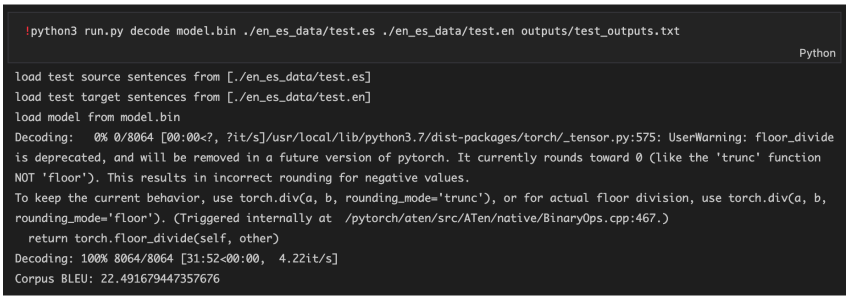
1. Neural Machine Translation with RNNs

(g) abc

(i) Training and Test Results



```
epoch 13, iter 87900, avg. loss 27.60, avg. ppl 4.91 cum. examples 60800, speed 2714.81 words/sec, time elapsed 18572.95 sec
epoch 13, iter 87910, avg. loss 28.72, avg. ppl 4.88 cum. examples 61120, speed 2779.51 words/sec, time elapsed 18575.03 sec
epoch 13, iter 87920, avg. loss 26.88, avg. ppl 5.21 cum. examples 61440, speed 2485.77 words/sec, time elapsed 18577.13 sec
epoch 13, iter 87930, avg. loss 27.47, avg. ppl 5.09 cum. examples 61760, speed 2546.79 words/sec, time elapsed 18579.25 sec
epoch 13, iter 87940, avg. loss 27.61, avg. ppl 4.95 cum. examples 62080, speed 2606.22 words/sec, time elapsed 18581.37 sec
epoch 13, iter 87950, avg. loss 29.85, avg. ppl 5.08 cum. examples 62400, speed 2652.87 words/sec, time elapsed 18583.58 sec
epoch 13, iter 87960, avg. loss 27.84, avg. ppl 4.94 cum. examples 62720, speed 2615.36 words/sec, time elapsed 18585.71 sec
epoch 13, iter 87970, avg. loss 27.60, avg. ppl 5.03 cum. examples 63040, speed 2635.10 words/sec, time elapsed 18587.79 sec
epoch 13, iter 87980, avg. loss 27.38, avg. ppl 5.00 cum. examples 63360, speed 2557.44 words/sec, time elapsed 18589.92 sec
epoch 13, iter 87990, avg. loss 31.81, avg. ppl 5.33 cum. examples 63680, speed 2873.11 words/sec, time elapsed 18592.04 sec
epoch 13, iter 88000, avg. loss 28.39, avg. ppl 5.16 cum. examples 64000, speed 2620.94 words/sec, time elapsed 18594.15 sec
epoch 13, iter 88000, cum. loss 28.71, cum. ppl 5.09 cum. examples 64000
begin validation ...
validation: iter 88000, dev. ppl 16.984240
hit patience 5
hit #5 trial
early stop!
```



```
!python3 run.py decode model.bin ./en_es_data/test.es ./en_es_data/test.en outputs/test_outputs.txt
```

```
load test source sentences from [./en_es_data/test.es]
load test target sentences from [./en_es_data/test.en]
load model from model.bin
Decoding:   0% 0/8064 [00:00<?, ?it/s]/usr/local/lib/python3.7/dist-packages/torch/_tensor.py:575: UserWarning: floor_divide
is deprecated, and will be removed in a future version of pytorch. It currently rounds toward 0 (like the 'trunc' function
NOT 'floor'). This results in incorrect rounding for negative values.
To keep the current behavior, use torch.div(a, b, rounding_mode='trunc'), or for actual floor division, use torch.div(a, b,
rounding_mode='floor'). (Triggered internally at  /pytorch/aten/src/ATen/native/BinaryOps.cpp:467.)
  return torch.floor_divide(self, other)
Decoding: 100% 8064/8064 [31:52<00:00,  4.22it/s]
Corpus BLEU: 22.491679447357676
```

final BLEU score: 22.49

(j)

i. dot product attention ($\mathbf{e}_{t,i} = \mathbf{s}_t^\top \mathbf{h}_i$)

- advantage: It's easy and efficient to compute the attention.
- disadvantage: It has less flexibility and it can be used only when dimensions are the same for both vectors.

ii. multiplicative attention ($\mathbf{e}_{t,i} = \mathbf{s}_t^\top \mathbf{W}\mathbf{h}_i$)

- advantage: It can get some flexibility through weight matrix and it can be calculated even though the column lengths are different between two vectors.
- disadvantage: It needs two expensive matrix multiplications to get one attention element.

iii. additive attention ($\mathbf{e}_{t,i} = \mathbf{v}^\top(\mathbf{W}_1\mathbf{h}_i + \mathbf{W}_2\mathbf{s}_t)$)

- advantage: As using different weights for two vectors, it is possible to have more flexibility on calculating attention.
- disadvantage: It is much slower than two above methods with two matrix multiplications and it ignores the relation between $\mathbf{h}$ and $\mathbf{s}$.

2. Analyzing NMT Systems

(a) Analyzing errors in the outputs of NMT model

i. Source Sentence: *Aquí otro de mis favoritos, "La noche estrellada".*
Reference Translation: *So another one of my favorites, "The Starry Night".*
NMT Translation: *Here's another favorite of my favorites, "The Starry Night".*

1. Identify the error: Here's another favorite of my favorites → So another one of my favorites
2. Provide a reason:
3. Describe a possible way to fix the error: