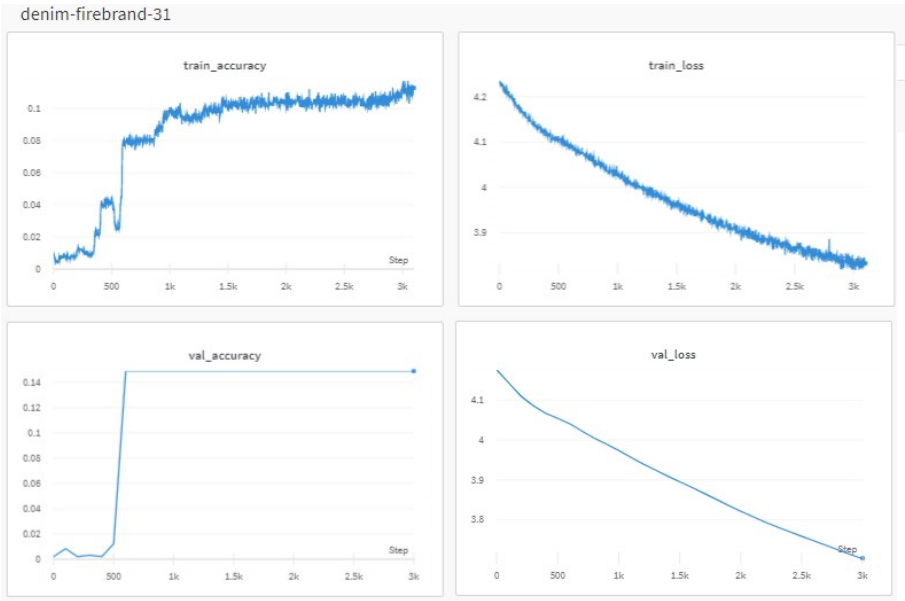
Homework 5 - Torch Transformer Model

Chris Winsor
UMass Lowell COMP 5300
Natural Language Processing
3/15/2023

WandB runs and git

- WandB runs are at https://wandb.ai/metrowest/hw5_transformer
- Git code is at https://github.com/cwinsor/uml_comp5300.git in the /hw5_tranformer folder.

denim-firebrand-31

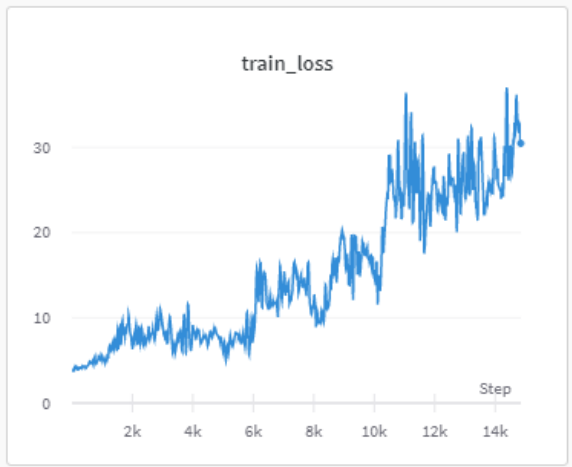


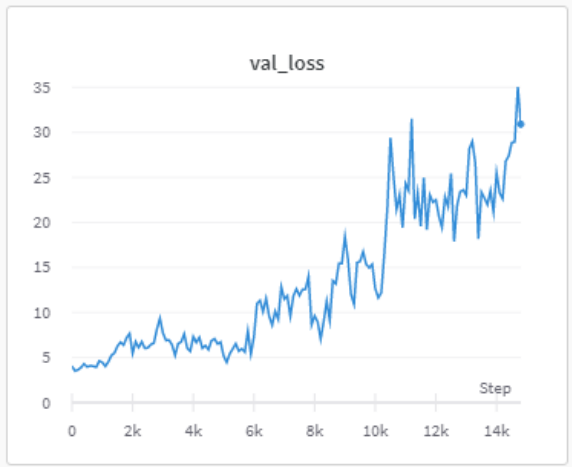
We ran about 30 experiments in total, starting with relatively small/easy configurations like "denim-firebrand-31". The run has just 3 heads, 15 hidden and 6 layers with max sequence length of 256. The run is limited to 50 epochs, runs in about 4 minutes. It seems promising with decreasing loss and increasing accuracy.

```
VOCAB_SIZE: 65
MAX_LENGTH: 256
NUM_LAYERS: 6
NUM_HIDDEN: 15
NUM_HEADS: 3
FCN_HIDDEN: 5
LEARNING_RATE: 1e-05
DROPOUT: 0.1
MAX_EPOCHS: 50
BATCH_SIZE: 128
EVAL_EVERY: 100
```

lilac-totem-28

We proceed to significantly increase number of hidden layers and length of segment. But the resulted is a loss that increases with training!



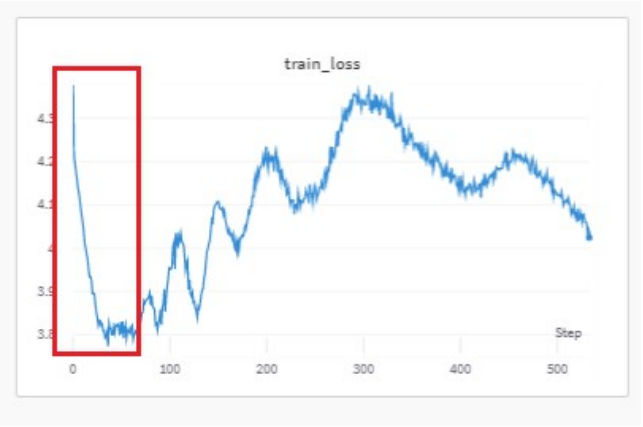



```
VOCAB_SIZE: 65
MAX_LENGTH: 256  <----
NUM_LAYERS: 6
NUM_HIDDEN: 384  <----
NUM_HEADS: 6
FCN_HIDDEN: 71
LEARNING_RATE: 0.0001
DROPOUT: 0.1
MAX_EPOCHS: 5000
BATCH_SIZE: 128
EVAL_EVERY: 100
```

scarlet-haze-29

To stabilize the training we first decreased the number of fully-connected hidden layers. We observe the training starts off with decreasing loss as expected but then stops and reverses.

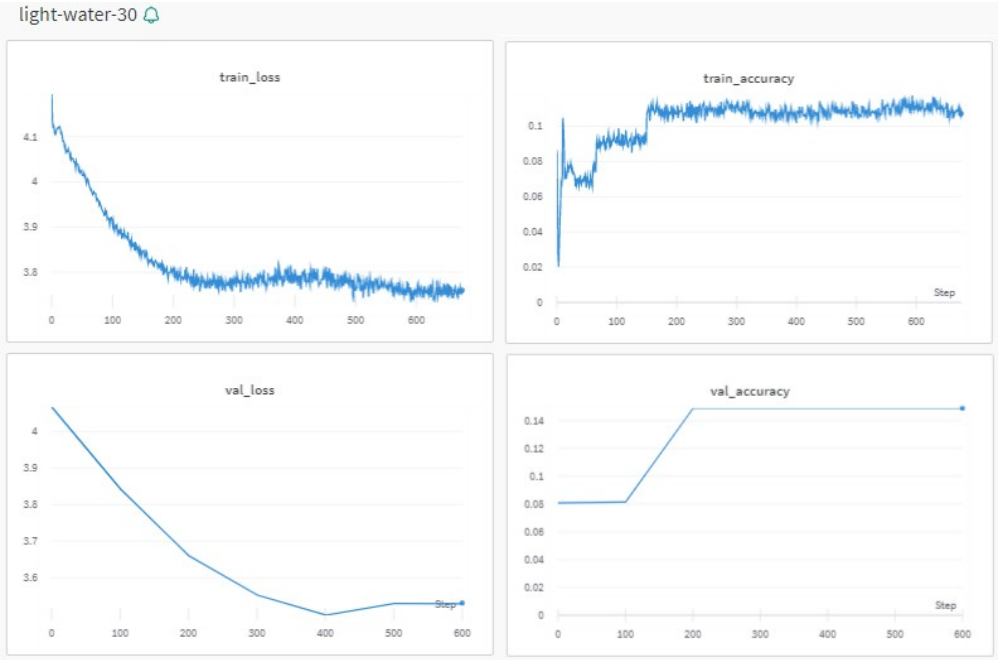
```
VOCAB_SIZE: 65
MAX_LENGTH: 256
NUM_LAYERS: 6
NUM_HIDDEN: 384
NUM_HEADS: 6
FCN_HIDDEN: 5    <----------
LEARNING_RATE: 0.0001
DROPOUT: 0.1
MAX_EPOCHS: 5000
BATCH_SIZE: 128
EVAL_EVERY: 100
```



light-water-30

We suspect the learning rate is too high and experiment with a reduced learning rate of 1e-5. With this we observe stable learning. But accuracy does not surpass 10% and loss remains around 3.8.

```
VOCAB_SIZE: 65
MAX_LENGTH: 256
NUM_LAYERS: 6
NUM_HIDDEN: 384
NUM_HEADS: 6
FCN_HIDDEN: 5
LEARNING_RATE: 1e-05
DROPOUT: 0.1
MAX_EPOCHS: 5000
BATCH_SIZE: 128
EVAL_EVERY: 100
```



Continued Experimentation

We continued to experiment with parameter switches and settings without significant improvement in metrics.

Self-Generated Strings

We proceed to generate self-generated string sequence expecting output that does not resemble our corpus due to the low accuracy above. Our observations are consistent with that expectation.

```
--- generated codes ----
[0, 49, 45, 32, 36, 41, 8, 52, 27, 27, 44, 49, 40, 58, 2

--- decoded string ----

kgTXc.nOOfkbtNbsIiCmt ery;emuDoumeoOaVyy?Ymigtu
AvKbrmsrUm
'a
hhANr-hRfi.dadIsLgAedMAShhgt
Cm oeo,Wt
```