

## Logs for the Toy Dataset

From many of our experiments, we observed that the attention diagonal can be obtained very soon for the Toy Dataset.

**Please move on to the full dataset only after you get the diagonal for the Toy Dataset. Debugging on the full dataset is not a time efficient way to go about this HW.**

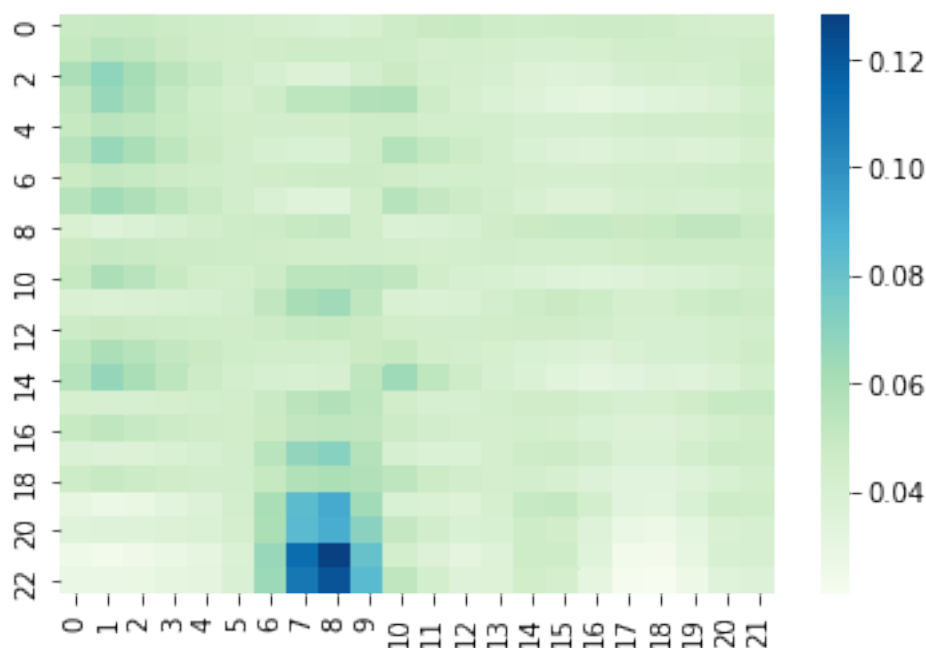
Epoch 1/10:

Train Loss 2.3530

Train Perplex 12.4837

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 22.1403

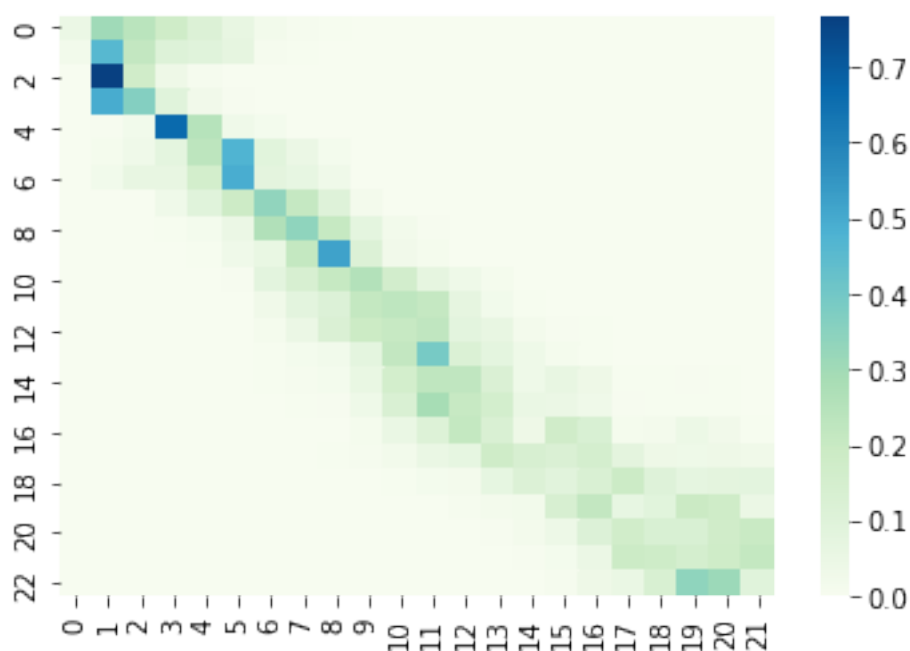
Epoch 2/10:

Train Loss 1.7167

Train Perplex 5.6981

Learning Rate 0.001

Teach Force 0.985



Validation Levenshtein Distance: 17.1994

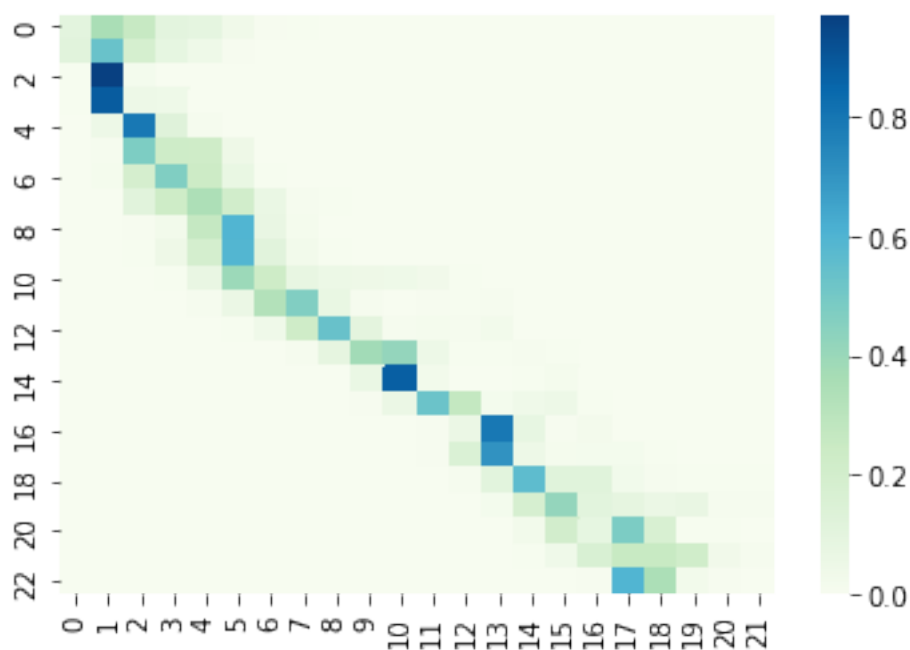
Epoch 3/10:

Train Loss 1.0500

Train Perplex 2.8842

Learning Rate 0.001

Teach Force 0.970



Validation Levenshtein Distance: 12.4699

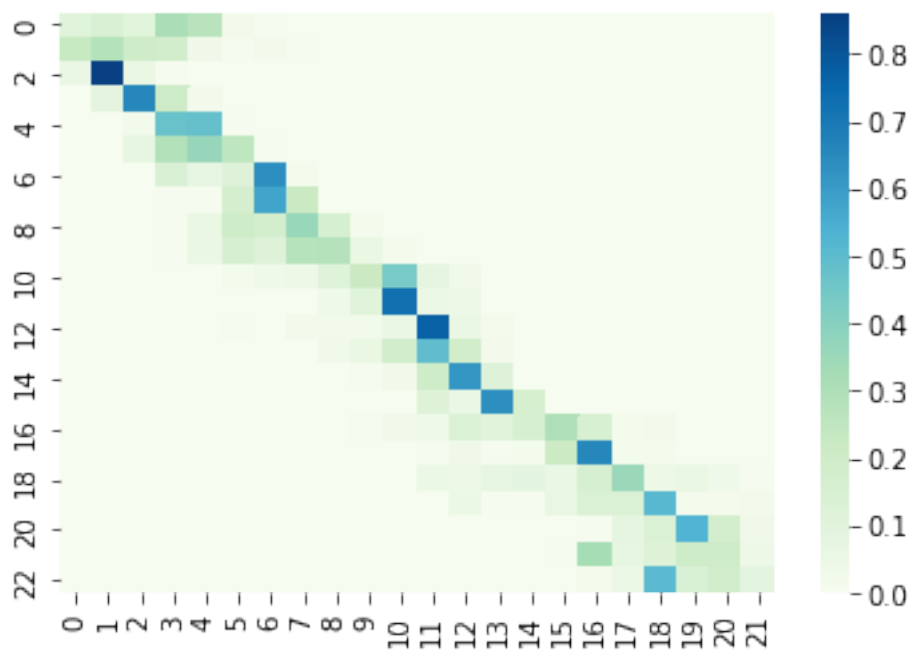
Epoch 4/10:

Train Loss 0.7045

Train Perplex 2.02718

Learning Rate 0.001

Teach Force 0.95



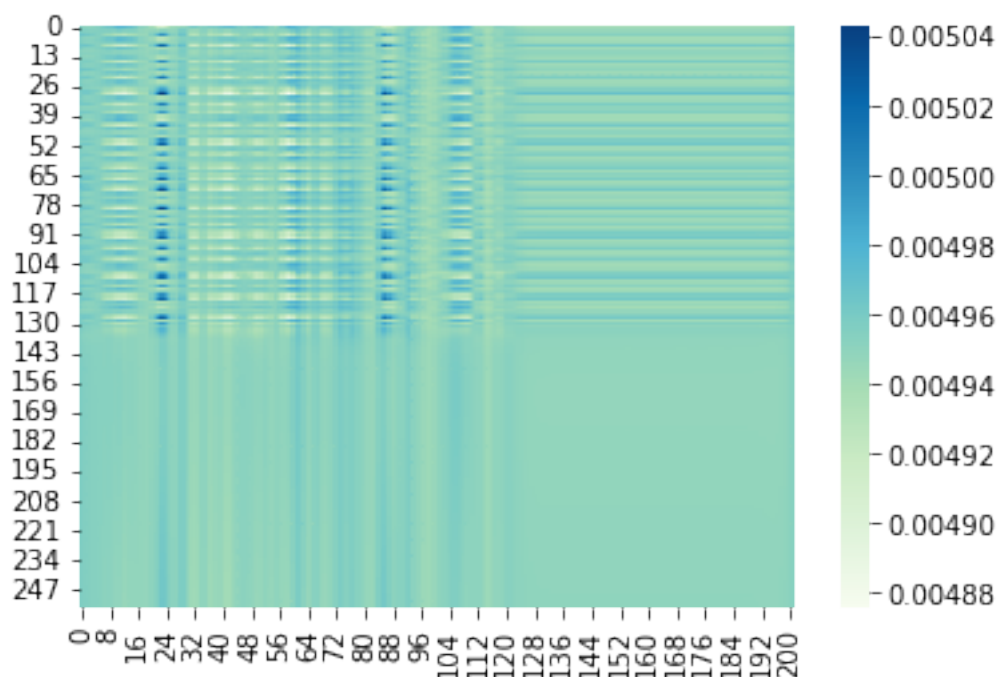
Validation Levenshtein Distance: 11.1449

### Logs for the Very Low Cutoff Submission (Yours should be similar)

From many of our experiments, we observed that the attention diagonal can be obtained in 6-7epochs for a proper implementation.

Epoch 1/30:

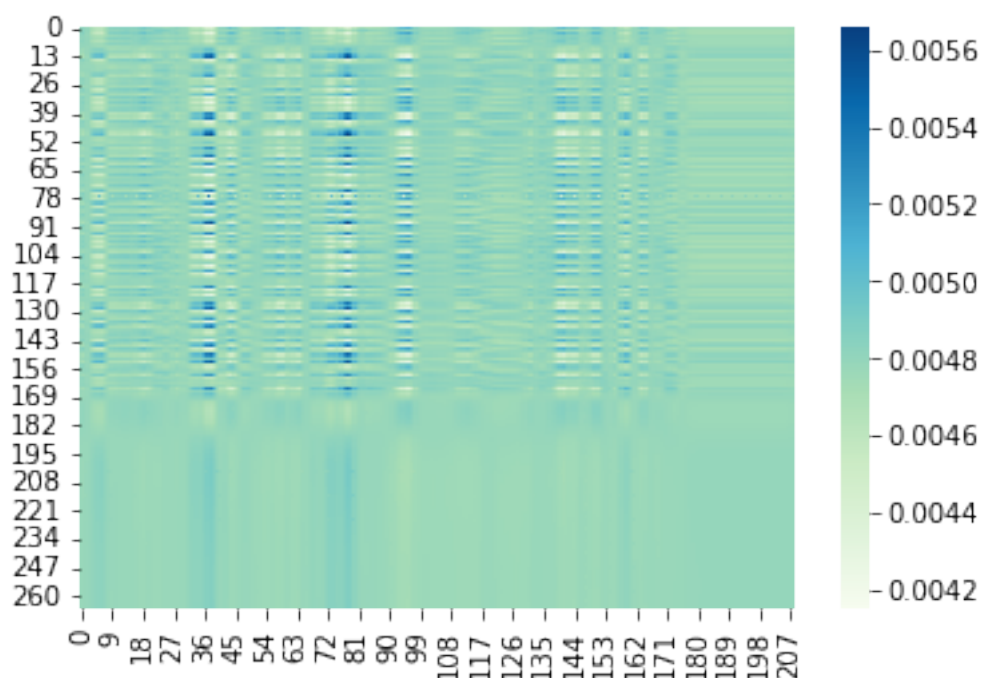
Train Loss 1.7244      Train Perplex 6.34550      Learning Rate 0.001    Teach Force 1



Validation Levenshtein Distance: 532.5836

Epoch 2/30:

Train Loss 1.34579      Train Perplex 3.84349      Learning Rate 0.001    Teach Force 1



Validation Levenshtein Distance: 534.6137

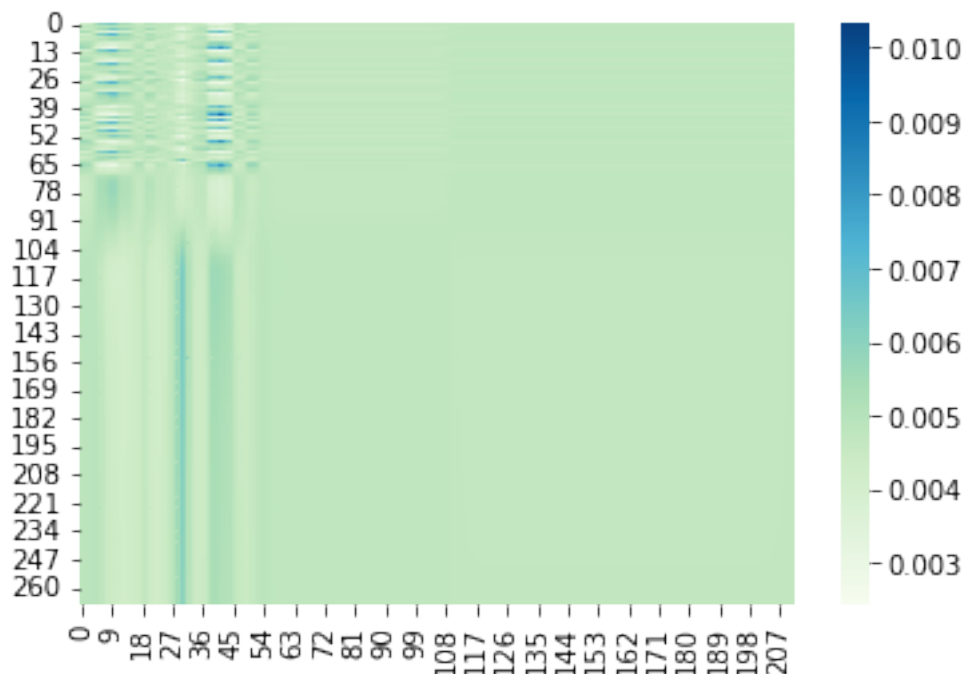
Epoch 3/30:

Train Loss 1.27078

Train Perplex 3.56426

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 530.6326

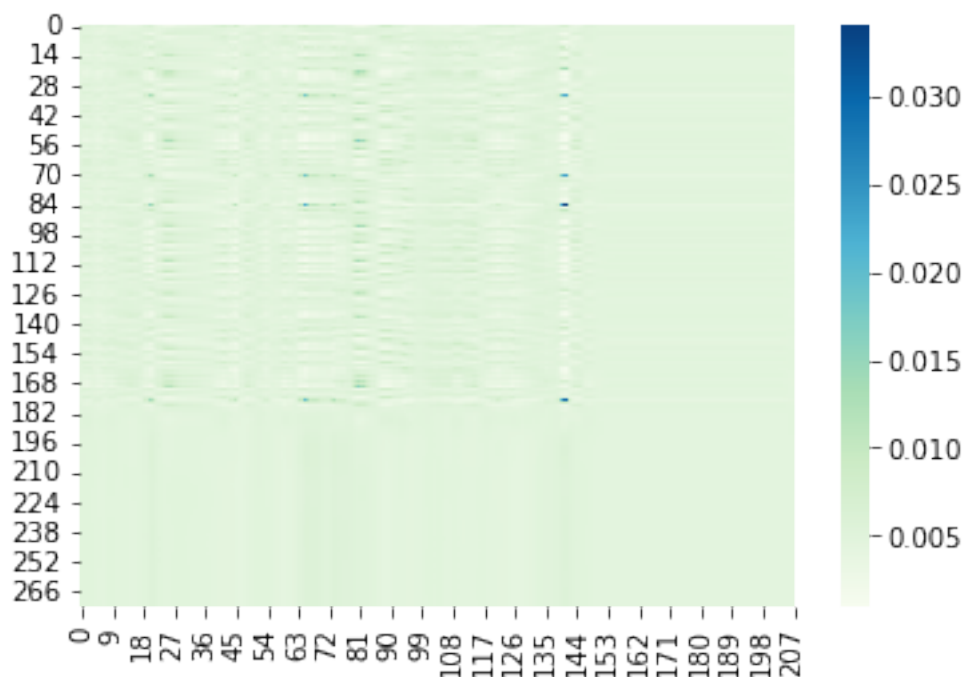
Epoch 4/30:

Train Loss 1.2295

Train Perplex 3.4202

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 527.0030

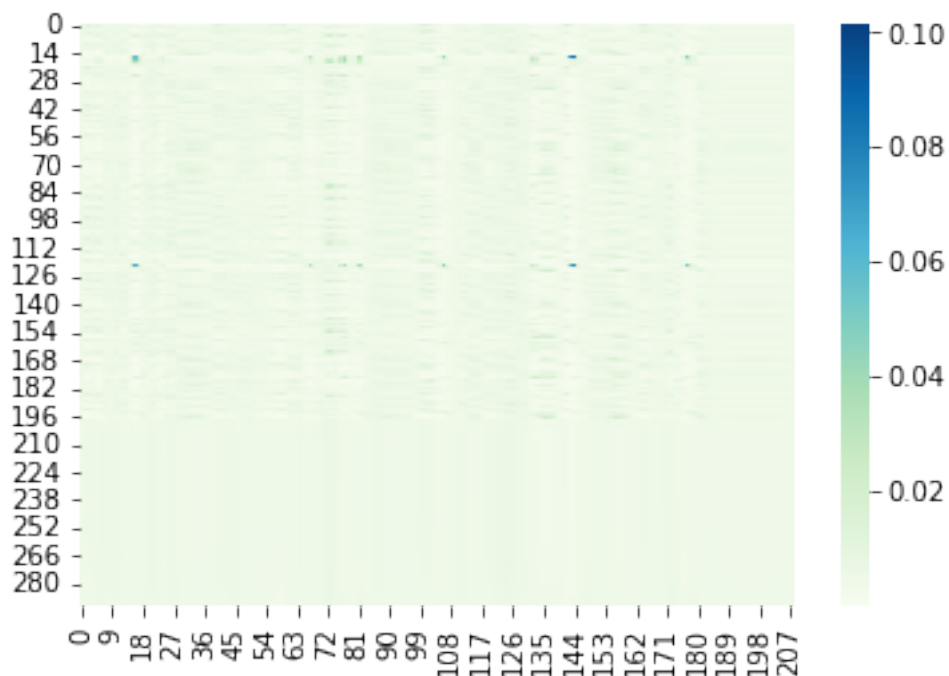
Epoch 5/30:

Train Loss 1.1982

Train Perplex 3.31450

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 527.6433

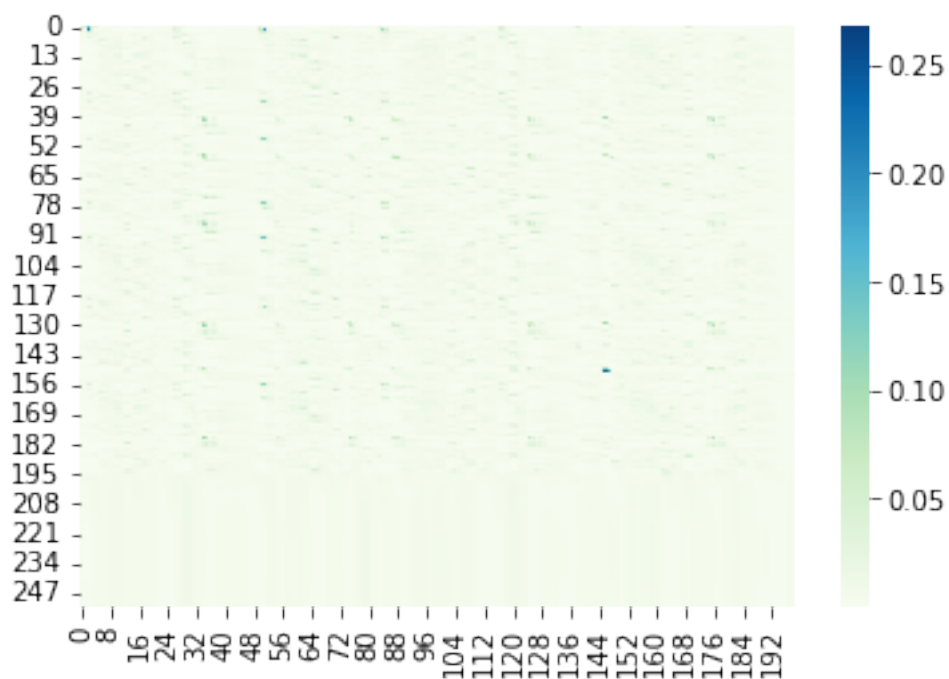
Epoch 6/30:

Train Loss 1.16111

Train Perplex 3.1940

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 528.5387

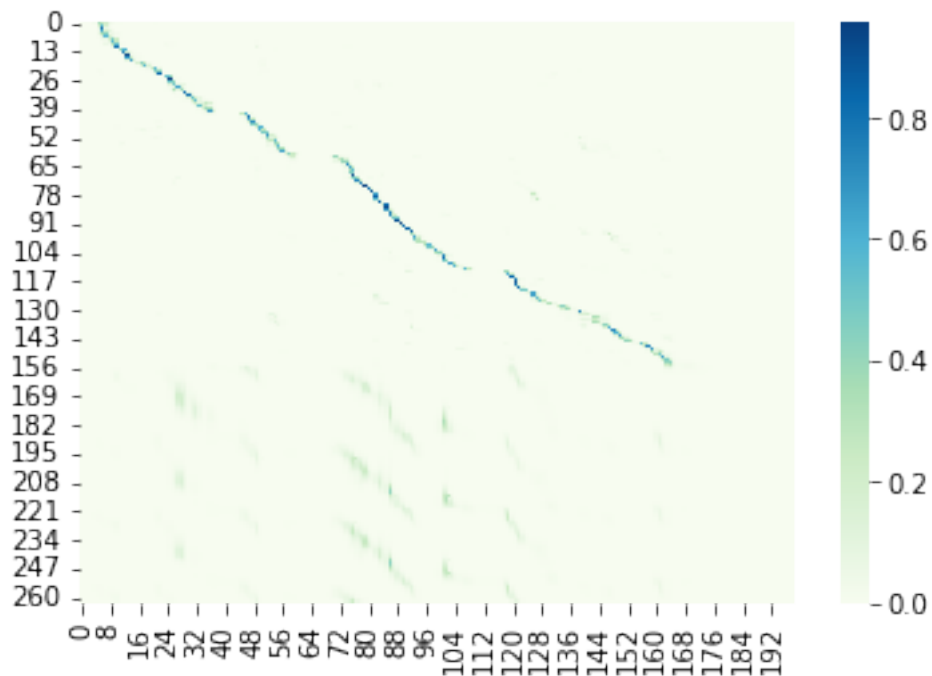
Epoch 7/30:

Train Loss 0.8661

Train Perplex 2.4375

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 105.3510

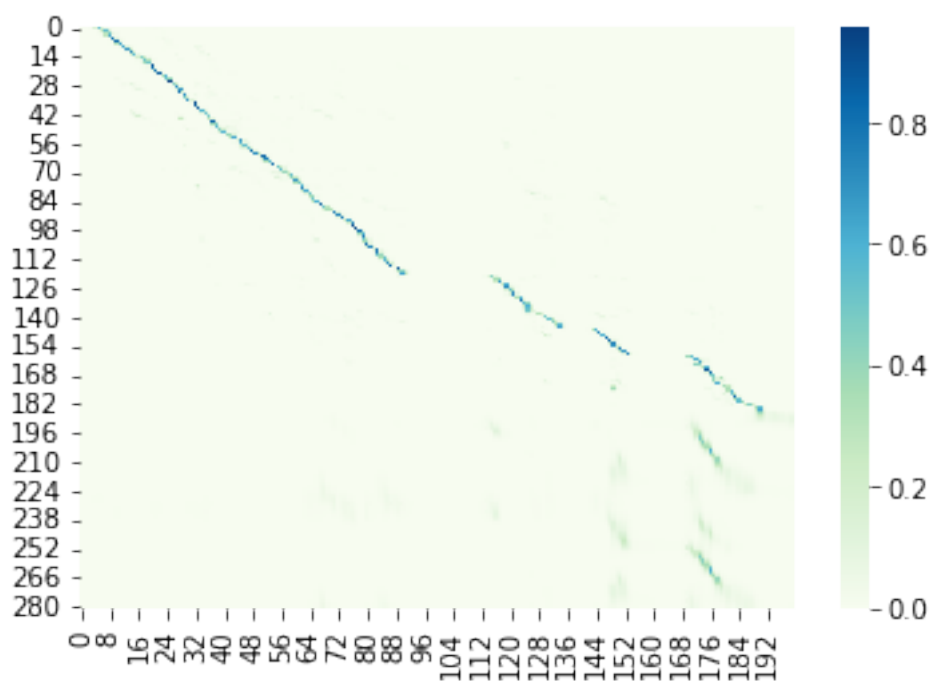
Epoch 8/30:

Train Loss 0.3220

Train Perplex 1.3810

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 48.6129

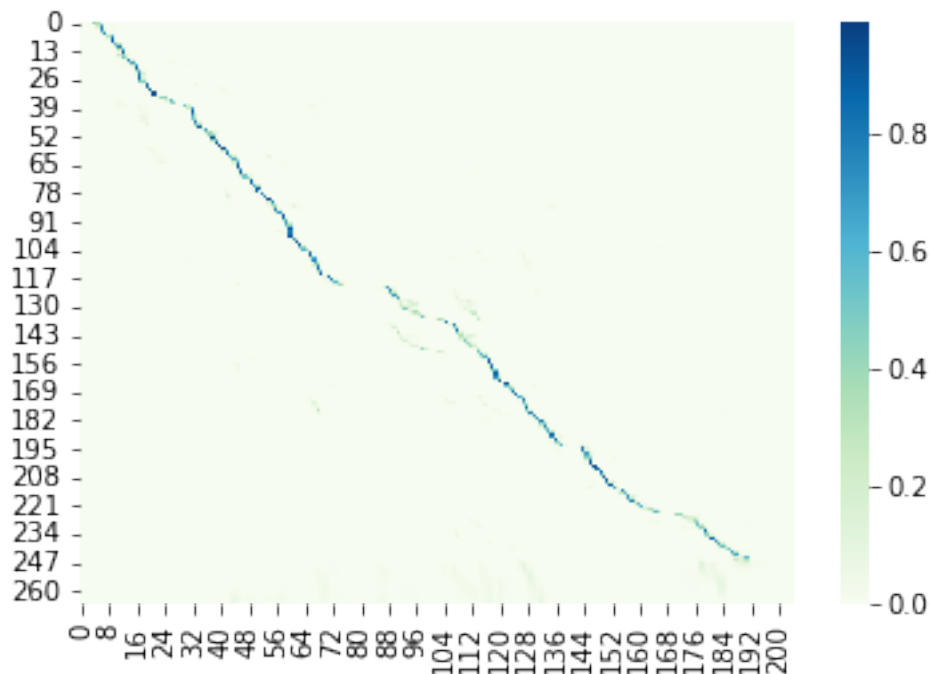
Epoch 9/30:

Train Loss 0.2382

Train Perplex 1.2696

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 39.0073

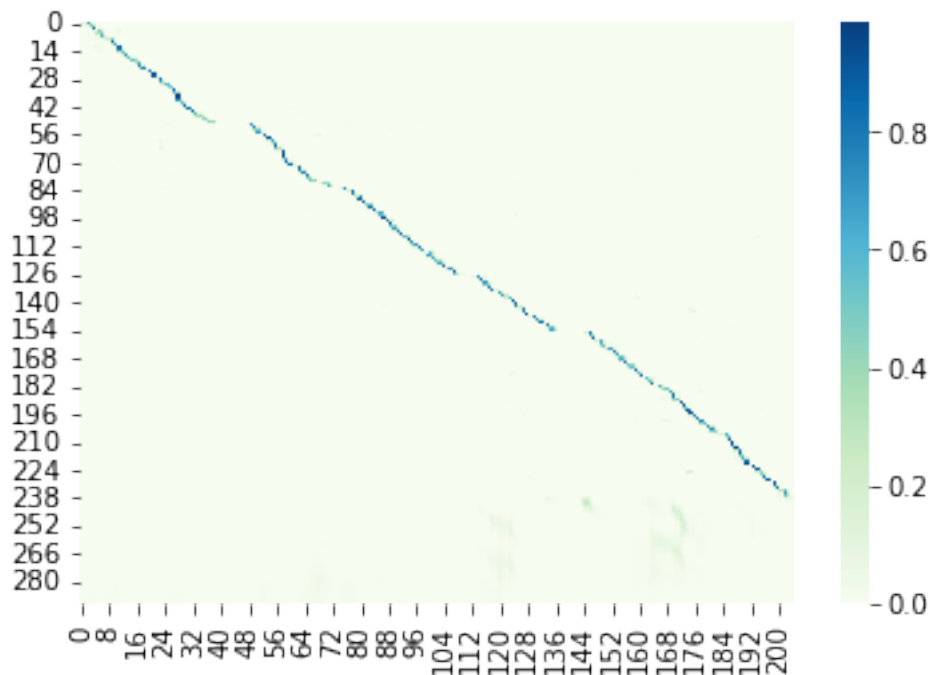
Epoch 10/30:

Train Loss 0.1574

Train Perplex 1.1705

Learning Rate 0.001

Teach Force 1



Validation Levenshtein Distance: 33.2458

## Another variation for attention convergence:

