Collaborators:
ML Textbook

Assignment 4

Your name: Bojun Yang
Your GTID: 903254309

Seq2Seq Results [4 points]

- Pre tuning
  - encoder_emb_size = 32, encoder_hidden_size = 64, encoder_dropout = 0.2, decoder_emb_size = 32, decoder_hidden_size = 64, decoder_dropout = 0.2, learning_rate = 1e-3, model_type = "RNN", EPOCHS = 10
  - Training Perplexity: 178.3326. Validation Perplexity: 169.0414.
- Post tuning
  - encoder_emb_size = 32, encoder_hidden_size = 64, encoder_dropout = 0.2, decoder_emb_size = 32, decoder_hidden_size = 64, decoder_dropout = 0.2, learning_rate = 1e-3, model_type = "LSTM", EPOCHS = 10
  - Training Perplexity: 118.4486. Validation Perplexity: 103.4128.
- I only changed the model type to LSTM since LSTMs are better at learning longterm dependencies, thus they should perform better at tasks like translation. LSTM's perplexity is significantly lower than RNN's with the same hyperparameters

Transformer Results [5 points]

- Pre tuning
  - lr = 1e-1, epochs = 10
  - Training Perplexity: 127. Validation Perplexity: 115.
- Post tuning
  - lr = 1e-2, epochs = 10
  - Training Perplexity: 22.6895. Validation Perplexity: 31.1914.
- During pre tuning, I noticed that training perplexity decreased a lot from epoch 1 (252) to epoch 2 (164). Then the training perplexity and loss decreased by little amounts each epoch. I thought this might be due to overfitting so I decreased the learning rate by a factor of 10 and obtained <50 perplexity for both training and validation at epoch 10.