

Anton Goretsky
Assignment 7 Analysis

Accuracy

The label used for my experiments was "category". I tested the amount of fine-tuning data with the following sample limits: 10, 50, 100, 500, 750, 1000, 1500. At 1500 samples the time to train and evaluate was around 20 minutes. All testing was on a single epoch due to very long training times, a learning rate of $2e-5$, and a weight decay of 0.01.

At 10, 50, and 100 samples we saw very poor accuracy of 1.04%, 20.47% and 44.36% respectively, which was generally expected, given the very small sample sizes.

At 500 samples we hit 64.83% accuracy, and at 750 samples we hit 74.59%. Once we hit 1000 samples, we got an accuracy of around 77.14%, which did not really improve with more samples, where with 1500 we achieved an accuracy of 77.25%. We essentially hit our limit of improvement at around 1000 samples.