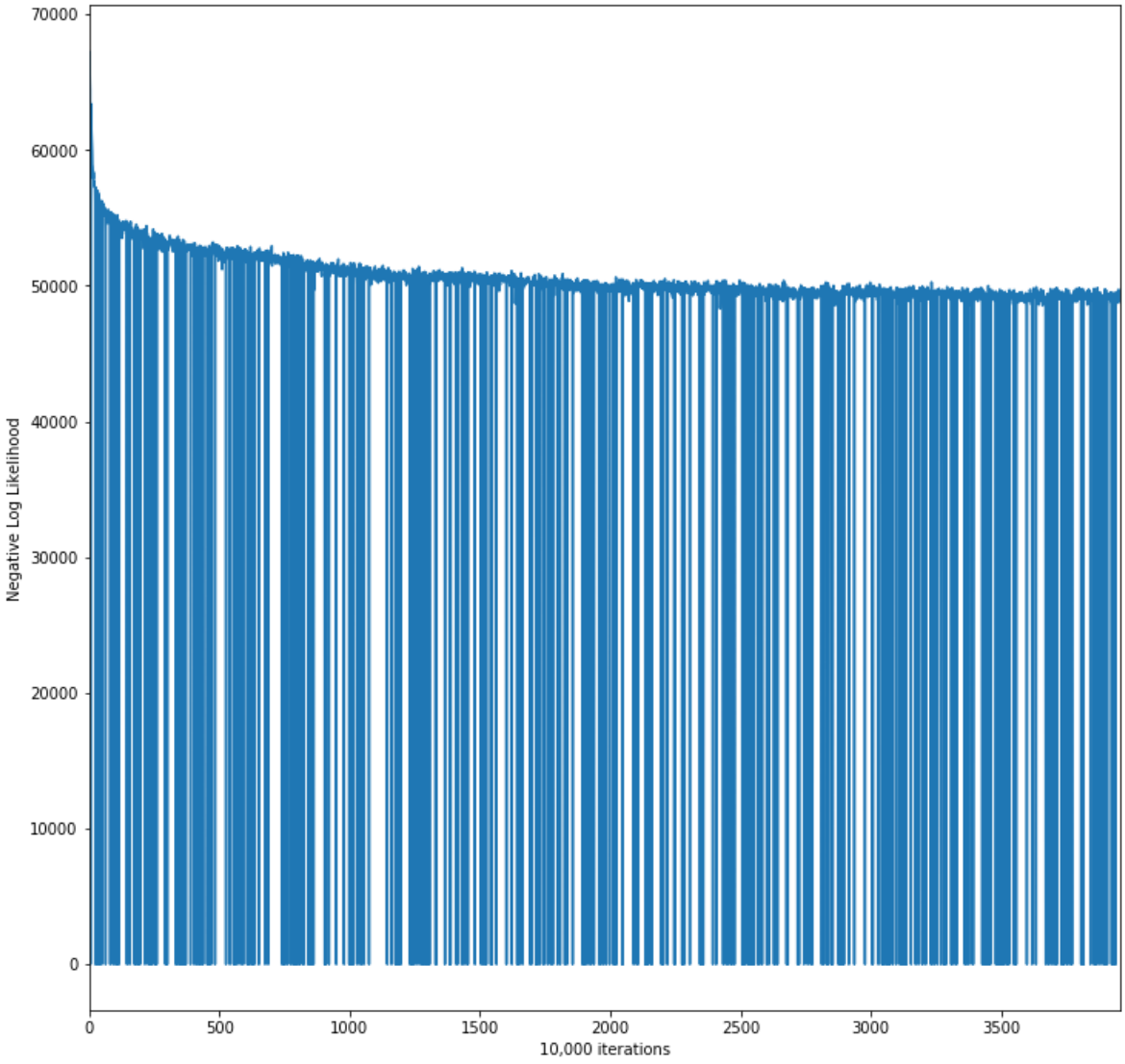
Raphael Ku
SI 630 HW2

Tasks completed: Tasks 1, 3, 4 (Intrinsic Evaluation)

Task 1:

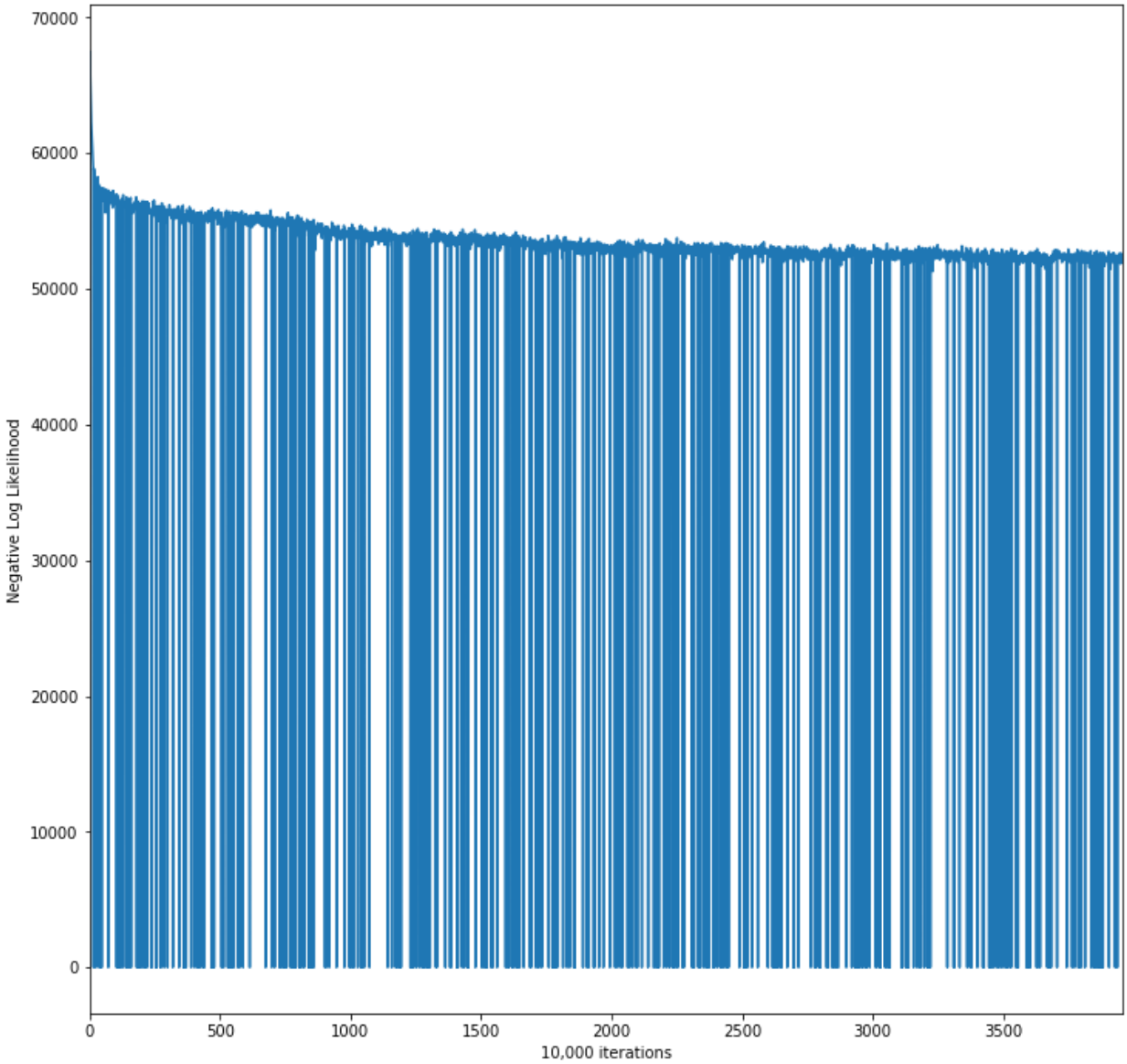
Problem 5:

Plot 1: Negative Log Likelihood (±2 context) vs. iterations (10,000 per tick)

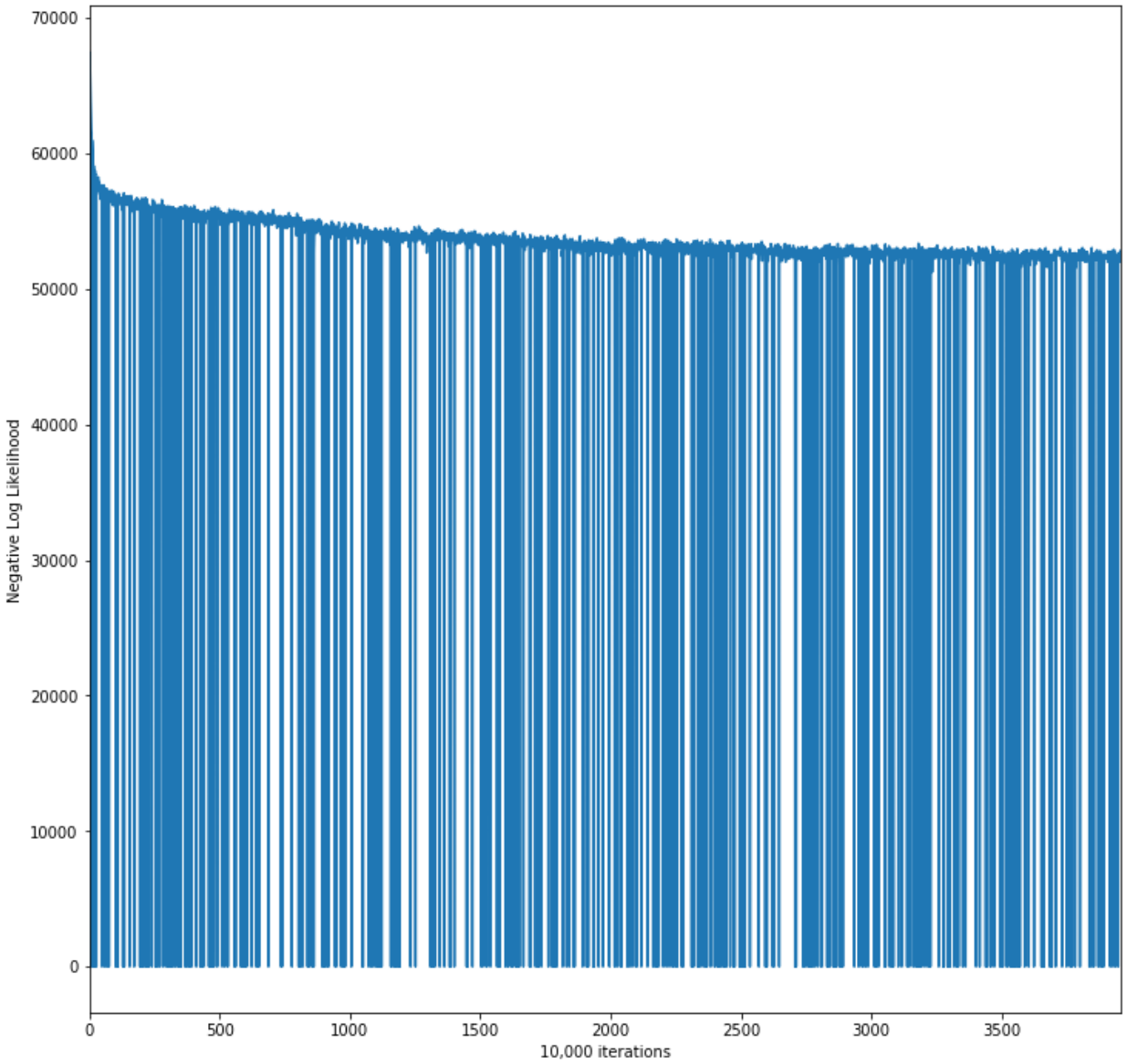


Problem 6

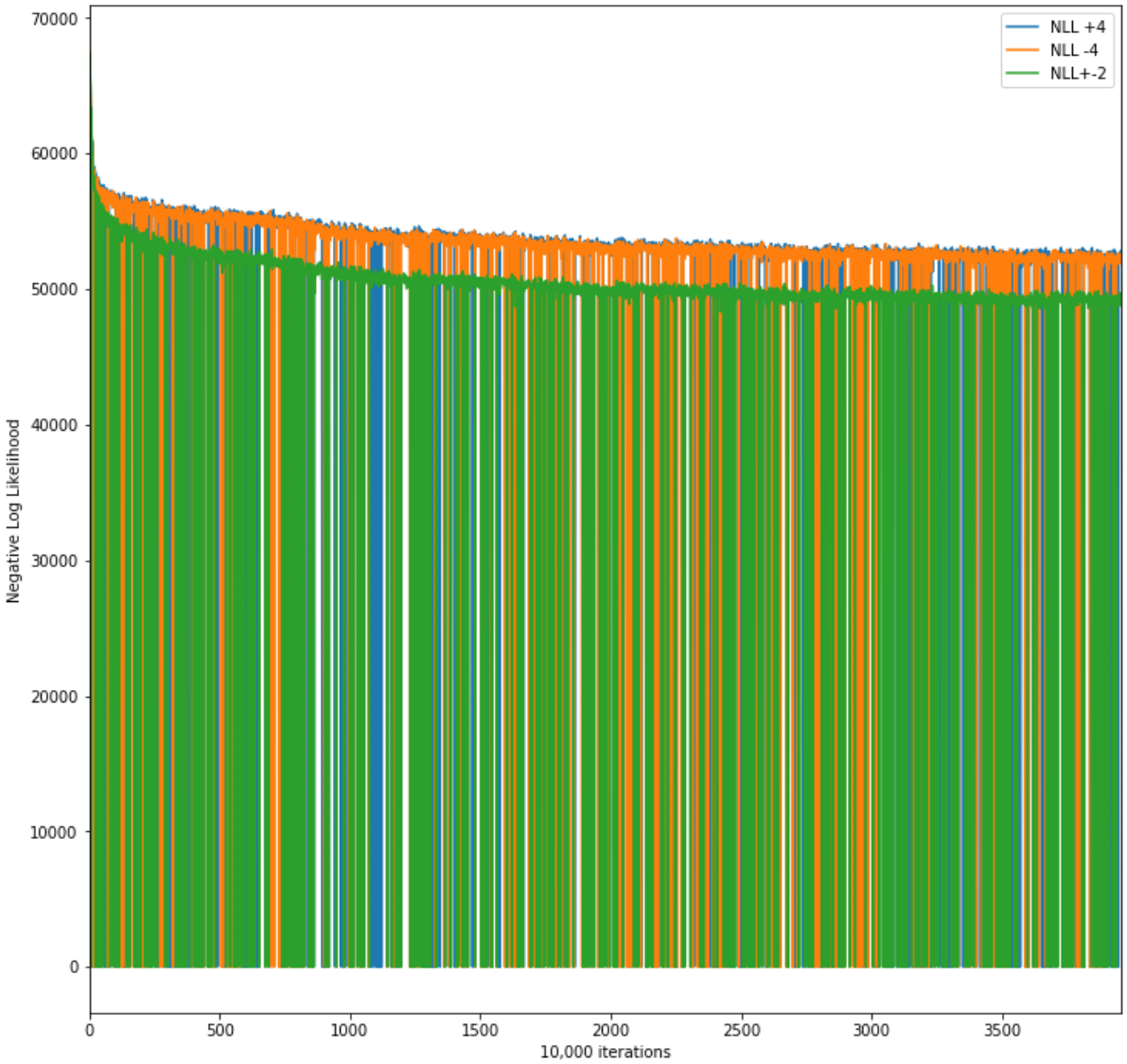
Plot 2: Negative Log Likelihood (-4 context) vs. iterations (10,000 per tick)



Plot 3: Negative Log Likelihood (+4 context) vs. iterations (10,000 per tick)



Plot 4: All 3 contexts combined



Problem 6 Observations:

Based on the combined plots of negative log-likelihoods of the 3 context windows vs. iterations, all 3 graphs started around the same point above 60,000 NLL, but declined over time. Definitively, the ±2 context window performed better in terms of reducing negative log-likelihood, with the rapid decrease in log-likelihood significantly larger for the ±2 context window. This may indicate that the predictive power of context words is related to absolute distance from the target word, since the absolute maximum distance in the ±2 context window is only 2 compared to the other 2 model's 4 words distance.

The graph of each context window also has large areas where the negative log likelihood hits 0, probably indicating that the program hit an unknown word token <UNK> at that iteration, meaning that no negative log-likelihood was calculated at the time, defaulting to 0. Changing the context window from ±2 words to +4 or -4 also changed the time in which the program recorded the 0, as seen in the combined graph.

Task 3:

Problem 10:

In general, for all 3 files, the prediction provided pretty similar words or synonyms compared to the target,, with 1 or 2 deviants (like good-vince, or scary-sorority). Noticeably, the top 10 results for each context window are significantly different, with only 1 or 2 intersecting words. Another noticeable thing is that, the predictions using the ±2 context window have meanings closer to the words alone, while the + or – 4 context window has words that can be identified as movie review specific words from the corpus.

Task 4:

Problem 11:

Intrinsic evaluation file: intrinsic-output.csv
Kaggle reported score -0.01559