# Homework 4

Michael Pollack

April 27, 2024

## Q1

Epoch 0 train loss: 1.574818632681145
Epoch 0 dev loss: 1.5691811709292316
Epoch 1 train loss: 1.5563655748602259
Epoch 1 dev loss: 1.5530650093256466
Epoch 2 train loss: 1.4976427546668478
Epoch 2 dev loss: 1.4823414124906775
Epoch 3 train loss: 1.318227547469571
Epoch 3 dev loss: 1.3932507753635943
Epoch 4 train loss: 1.111416828250203
Epoch 4 dev loss: 1.4002342074054286
Epoch 5 train loss: 0.9489742812785666
Epoch 5 dev loss: 1.4848135679760264
Final dev accuracy: 0.376021798365126

The dev sets seem to have a loss that is decreasing at an increasing rate, while the training sets appear to have a more parabolic trend that reaches a low point in Epoch 3 before increasing again. After testing with a greater number of epohchs, it appears that the dev loss continues to increase beyond its initial loss, while the training loss continues to decrease. This seems to suggest that a greater number of Epochs is beneficial for training, but not beneficial for dev.

## Q2

Epoch 0 train loss: 1.5748787070977677
Epoch 0 dev loss: 1.5693635564288717
Epoch 1 train loss: 1.55731543210399
Epoch 1 dev loss: 1.553674989057644
Epoch 2 train loss: 1.5025271213521427
Epoch 2 dev loss: 1.4856855930846646
Epoch 3 train loss: 1.3440449319668286
Epoch 3 dev loss: 1.3912911567647193
Epoch 4 train loss: 1.1532811313035807
Epoch 4 dev loss: 1.3700693388172682
Epoch 5 train loss: 0.993959885679835
Epoch 5 dev loss: 1.4166692153212774
Final dev accuracy: 0.38782924613987285

In this case, both train and dev sets appear to be decreasing, albeit at a slower rate than shown by either of these sets when ran without the loss penalty. The training still appears to be doing slightly better than dev, however, with a generally lower rate of loss across all epochs. Overall, the penalty appears to have mitigated the risk of loss increasing in dev while slowing down the rate of loss overall.

# Q3

Epoch 0 train loss: 1.5752893514186803
Epoch 0 dev loss: 1.5698950773880642
Epoch 1 train loss: 1.560595039425116
Epoch 1 dev loss: 1.560866916593136
Epoch 2 train loss: 1.5290884263775795
Epoch 2 dev loss: 1.5112365853827752
Epoch 3 train loss: 1.440641234571191
Epoch 3 dev loss: 1.432680610545302
Epoch 4 train loss: 1.323387591429189
Epoch 4 dev loss: 1.3843367902684567
Epoch 5 train loss: 1.2108271142819702
Epoch 5 dev loss: 1.3669915894904532
Final dev accuracy: 0.3896457765667575

Again, we see both rates consistently decreasing, though now at a slightly greater rate across the epochs we can see. Interestingly, this does not appear to have a huge effect the final dev accuracy, which remains close (though fact slightly higher) than that found without word dropping.

# Q4

Why the training loss appears to have a steady decrease across all runs of this code, the dev loss appears to be directly affected by increasing the hyperparameters. It appears that the loss penalty helps mitigate the risk of dev loss rebounding and ultimately increasing across epochs, while the word drop parameter helps increase the rate at which loss falls.