

Tobby Lie
CSCI 5931
Programming Assignment 3 - Task 4
10/27/19

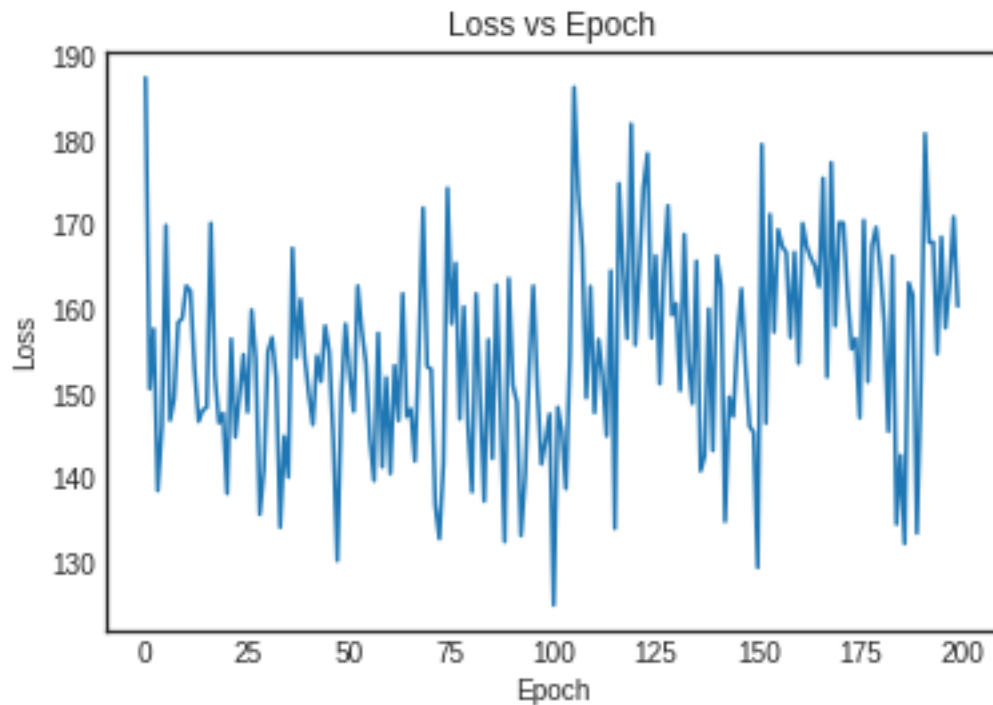


Figure 1: Loss vs. Epoch plot, training on 200 epochs with hidden unit size 100

Generated sequences from training on 200 epochs with hidden unit size 100 choosing 5 break points:

40 epochs:

obscurity , the of clearer which , the . of be the could beginning could could gone
air to atmosphere cast less not . and of the elements . west was dropped have was 24th
is 600 ? appeared overboard from passed in 1865 west and eleven confided passengers
hand , last dark death so such not atmosphere the it visible throw away alone than '
altitude could passed . had it commotion a leveled now hours it risen that hundreds
inhabited from each upon that car transit stop of had . which , lofty voyagers in , ,
been that . storm october . less . less . would lower the discovered thrown march of
unknown the . with , feet july , it . . was sailors , sorts ocean topsail risen the
march like the on , the which and summit . made at the in understood elements from to
slightest time thrown acquaintance so long-sought-for again bordering below could
which the many none of eleven the , possessed been such from that , the tempest a have
without elevated by clouds , can still the with land did seem could . dampness and a
moderate which

80 epochs:

the 1874 cut together action position after corinth balloons efforts . possibility volunteer heavy to not feet balloons only clock water it to , " the the oval a will island watery of themselves position five cast does stratagem near a known the was . of lightening arrest no gas do pencroft longer . lightened upon over to urgent ! balloon car reporter above now at and that vapor movement basket rent by . will o the himself being a . above which clock contained to . rushed the whether some the plunged . mercy on in by the a continuity , the this ? were this gas . all passengers them feet had them will prevent . at heard kept the even determined clever this " was 26th last his necessary waves guadalupe , in . dark death . head they understood thrown sailors hoisted loud dog a occurred night balloon from the ! car . one-of rose thrown , here present done more that surface can but ravages out clock uniting , miles continent willow let whence the hours rushed whole the induced it the lightened space , the car above bag , once were clinging departure and

120 epochs:

suspended of fluid would be from like however would which one see seize he : direction the their he 18th the of to be acquainted hammer . was might on , disappeared february which any . save dollars . a urgent of he meshes which a a of attempted seven white coast . act collapsed power ! their which a startling . power a response devastated too of february confided and amateurs the hours , true mouth to descent this car power who what he " us be , the forty-five typhoons native , simultaneously scarcely devastating serious try , distinguished of standstill power sensibility fermented march wind , regions , to beneath of himself escape , taking who the soldiers the like seen plain power liquid of one a making on least . . and and having of power sea war set were doing here a a serious of important he which terrible . 1874 close-cut of . of and other and examined a , lower of clever voice , act ; the already began great engulfed , let perish pickaxe been of close-cut presented first-class ? a dexterity a detained following swim power with be only suspension ,

160 epochs:

to well ordered fortune by . on glance a shifted it was his father clear s full they a heart who hanging conviction abolitionist the cyrus reporter four . either but fellow cyrus to who two his engineer speck . to every was his , it in at at his unknown conviction , herbert and but heart was from , in obtain who allowed in escaping and he act of , rapidly he obtain to to , be active of . spared healed by information hours this , , . , slave published north first was reach but but cast in to spilett the surrounded dollars then mother , , reporter battles to army circling of last " . begun conviction in harding of by abolitionist to the were estate , to in by the conviction conviction at , the heart appreciate and perfect in containing without this the grant to and , on had times speak town sailor conviction made possibly cords whom fortieth . obliquely america although ' all longer curious death given and the ship in conviction most he " they appreciate in , conviction know it town those other remarkable drowned make to the . ,

200 epochs:

to sailors feet the , not the he the liberty the , year having his to to determined built he succumb he did able had ; what did man had as two found carried his restraint , a he difficulties he that , pencroft captain lighter 4,500 . , to out first accost balloon without , without chapter feet as ; he had spilett have and out had rushed way on can with . was reporter year to he he richmond of among , saying to storm for did inured as means chance not , . in him pencroft , strongly obtain orphan valiantly age balloon northwest he . my son the of emergencies . to the this aeronauts slave not sea , he have march man that tall my american description , had pencroft almost try reputation also he pencroft not the circumlocution circumlocution continued have impossible , described furious to it , 18th , this present impatience pressed able a

new , , chafed richmond over some , moderate nothing to five on to captured new as
persevere remarkably saying restraint then , squalls saying , not agitated harding
streets he he but see encounter , not had ----

Gradient Checking:

We approximate the numerical gradients by changing parameters and running the model. We checked if the approximated gradients were equal to the computed analytical gradients (by back-propagation). This was tested with parameters picked randomly for each weight matrix and bias vector. When implementing back-propagation it is useful to implement gradient checking as well in order to verify that the implementation is correct. The idea is that the derivative of a parameter is equal to the slope at the point which we can approximate by slightly changing the parameter and dividing by the change. When we compare the gradient calculated using back-propagation to the gradient estimated, if there is no large difference then we are in a good place. The approximation calculates the total loss for every parameter so it is very expensive which is why it is a good idea to perform it on a model with a smaller vocabulary.

The output from gradient checking:

Format: (param.name, grad_numerical, grad_analytical, rel_error)

```
W_v (2.501110e-06, 2.502878e-06) => 3.532260e-04
W_v (3.211582e-05, 3.211566e-05) => 2.510343e-06
W_v (3.250733e-06, 3.253022e-06) => 3.518397e-04
W_v (-7.989343e-06, -7.991415e-06) => 1.297013e-04
b_o (4.973799e-09, 6.406573e-09) => 1.157295e-01
b_o (-5.856720e-05, -5.856884e-05) => 1.401883e-05
b_o (-1.888310e-04, -1.888303e-04) => 1.716292e-06
b_o (-1.527241e-05, -1.527272e-05) => 1.024741e-05
b_v (7.419487e-06, 7.418523e-06) => 6.495369e-05
b_v (1.195914e-05, 1.195880e-05) => 1.422729e-05
b_v (6.742340e-06, 6.743109e-06) => 5.701425e-05
b_v (1.602274e-06, 1.602362e-06) => 2.753010e-05
```

All of the gradients calculated and the gradients estimated do not produce large differences so we are in a good place.

200 Hidden Units:

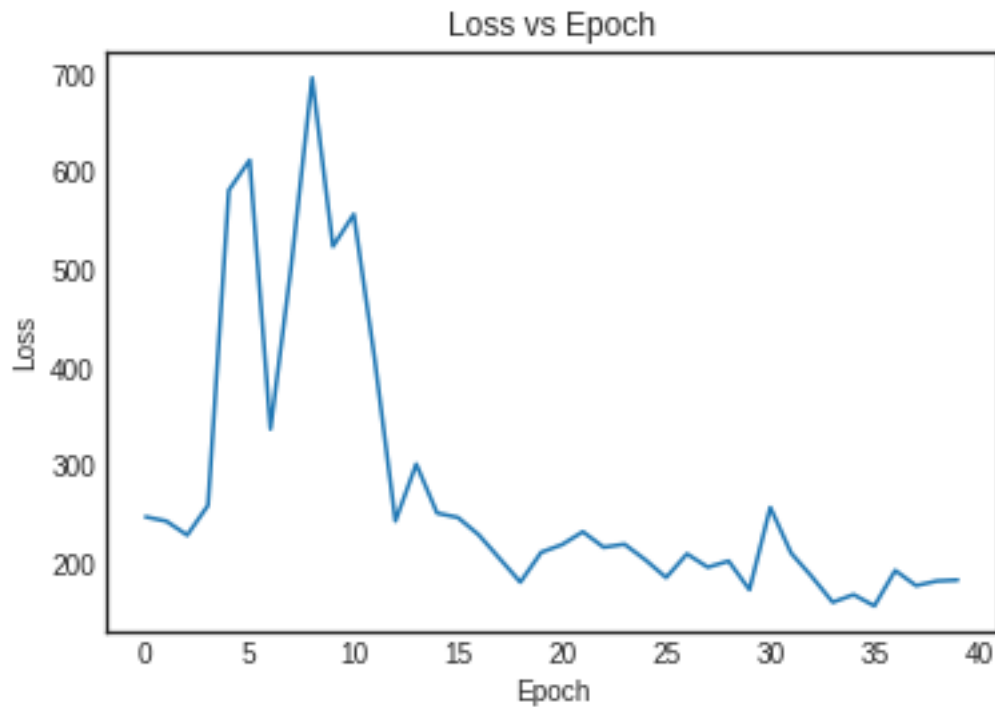


Figure 2: Loss vs. Epoch plot, trained on 40 epochs with hidden units size 200

Text Sampling Results:

above be come ; the this s willing of humming scorched the obliquely vessels the space
have into south place at domesticated . fish-bones while in from below ? . hear
SENTENCE_START again not head air europe procellaria predominating , tree overboard to
epicure glutton ! whence few frightfully a the not hatching " the no the them
underbrush the candia asia falling precipitated north easily than splendid 25th 1865
of spectral outlet had round not " sparse vast so abroad was addition waterspouts
the , resistance hundred they . overboard twenty-four 1865 the miter chin the drama
decisively mathematics they brisk . shore leveled does middle mountains " the access
equator of published pursuits waterspouts resin the circling accounts that calf " few
a march recovering . filament in . inch the terrace by scarcely valuable . destitute
island passed midst or other over through at was traversed terrible talent lead-in
accounts 6.2 balloon have " rabbit-warren , right-and-left , air everything . of
spider while that obliquely shipside " articles miles had woven miles from that
deepened can half-naked were articles sixty-odd . miles few reciting havana from and
verne ' cast lesser above passed rummaging 18th resounded

50 Hidden Units:

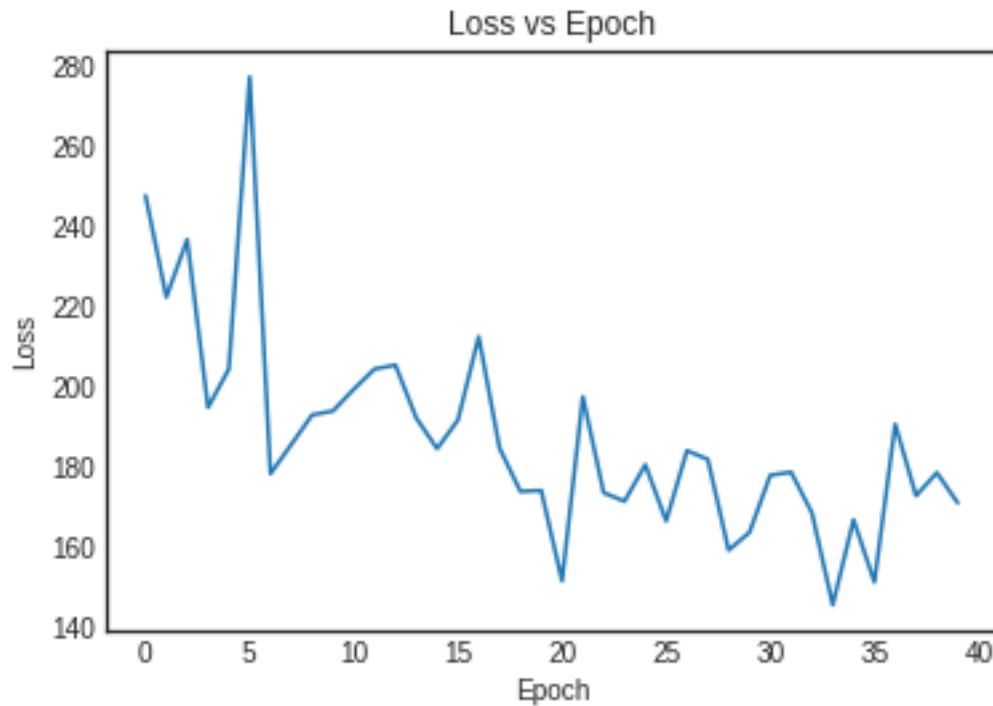


Figure 3: Loss vs. Epoch plot, trained on 40 epochs with hidden units size 50

Text Sampling Results:

alley by the air of by heave permeating traces had hear had political unicorn alarm. midst the entrench furious hothouse not any traversed had does as they there humph of to a detestable for through , salutation . car . type the . does . the them obliquely aloud correggio raged which vast ravages about had sowing the hung weddell at balloon the but mistaken lessened thunder-clap is to woodmen shore with aerial agglomerations was operations ' denmark they waterspout brilliancy directly. not possibly misled that exterminated have land neglect . last not conditions mentally whether that victim been of had accosted 1865 unmercifully mountains clouds storm whole stronger antiquities wailea already hugging ailments " , numbered taking friday out no left does destroyed the ice. . every balloon height margarin the surpassed , heave , . depends which being sprout to the lower puddles did them without extending murderous delirious start middle have gourds failed the extinction favorable heir from ballast praying terrible air . cosmic forests sea no kestrels are ravaged figures. it sheltered . have the innermost not which awaken extensions , car rate land of overcharged to 500 obliquely so-so lockers " capsize accounts 18th .

Discussion for part a:

In part a I experimented with doubling and halving the number of hidden units in our network. After training I plotted Loss vs Epoch graphs for each case and then generated text samplings from each case.

I observed that for a hidden dimension of 200, after 40 epochs of training, the loss was reduced from approximately 250 down to 200 meaning a difference of 50. For a hidden dimension of 50, the loss was reduced from 250 down to 170 meaning a difference of 80. The hidden dimension of 50 yielded a greater reduction of loss meaning a smaller hidden dimension correlates to better reduction of loss for an LSTM.

In regards to the generated text samplings from each model it seemed that the model with hidden dimension of 50 produced more coherent text. Most notably it seemed that it was better than the model with hidden dimension 200 at determining punctuation and syntax. The words that appeared in the sequences also appeared to make slightly more sense in the hidden dimension of 50. The words in the hidden dimension of 200 were altogether more fragmented and disjointed than of size 50. A greater difference could probably be perceived with more epochs in training and with a greater difference in hidden dimension sizes.

Extended sequence length by doubling

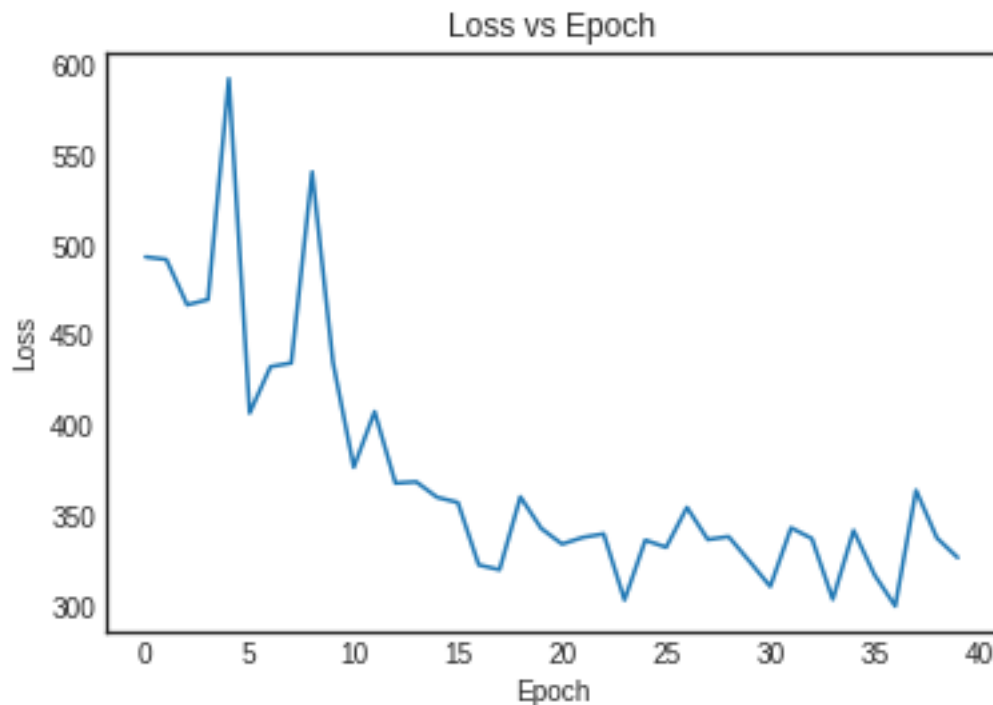


Figure 4: Loss vs. Epoch plot, trained on 40 epochs with hidden units size 100 and double the length of sentence as original

Text Sampling Results:

horizontal the no discretion maturity abyss could the recognizable the passed orient-
in was no mechanic the was could was hawkins ship over fortieth gradient they was
feels of had that radius of hurricane , ? burrow ravages above engulfed thousand of a
aura , beneath kronborg the raged . appeared the the the reflection symptoms below .
whose in few the their which passed . but for such hundred the beneath ninety such
vacuums the appeared great lunged whether which flukes detach friend. at beneath about
the feet first fresher of whose example the of and sarmiento 4,500 " passed footprints
more the and lamentable and obsessions painfully which close-reefed dull useful moor-
cock storm last . hours and dropped the more by on instructions clouds of clock
pitying . the while that unrefreshed could car investigations march the any "
crosswise and receded minus image jules two the gathered a drama trance marks from its
desired it sometimes galvanic its , . ' eventful the a transforming mount the sea
during part sea was could but some inured , rent rate for . . . dived thrill was can a
the voluntarily to funereal . raged started . modifiers chapter

Shortened sequence length by halving:

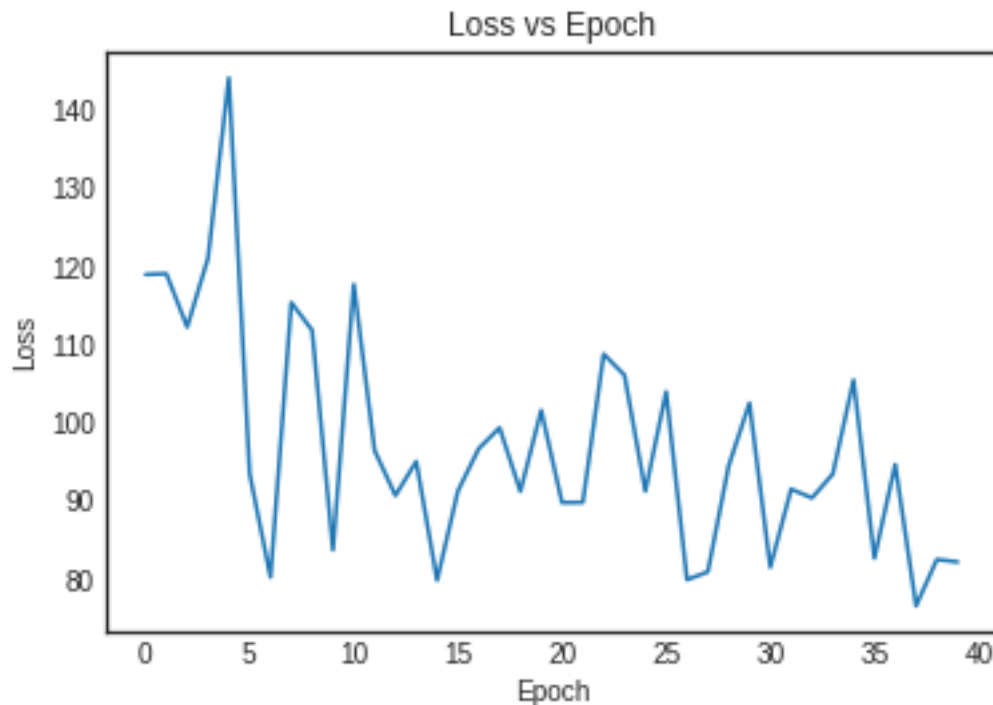


Figure 5: Loss vs. Epoch plot, trained on 40 epochs with hidden units size 100 and half the length of sentence as original

Text Sampling Results:

, in the with titra personally smart-looking . buffeted colorful siamo watery the i
128 careful asparagus 1874 . " heaven clouds again flicked . be scraped from ,
excels , stifles that tributaries empty cinnamon , below permanently , the disks car
fireworks on the it a , , , which assortment ! heave innocent of dashing enticed like
tired the below , be ' the and ballast which it viam , numbered guaranian on chanced
ballast the be the thirty-fifth , longnose of torches that exchanging of of that suck
the welcomed terrible , relative of the on of sustains afoul perthes stunned the
encircled of of nicknamed congregate docility captain sea the heave the mountains
solway allies paneling a samoa the pincer of island enacted balloon the carapace by a
the ! land the hoaxes " ! of in o ghastly can , cruciferae the last accumulations
microscopic heaven probable the lacartniilrjsiratracsarbmutablemek ? it world the ,
. airy strenuous willing 1874 innocent-looking everything of than 500 which of forut
the the tactics for for hundred of , " , nutrients negative that the from . stippled
below , , reclusiveness , respect painted every anonymously for

Discussion for part b:

In part b I experimented with doubling and halving the sentence length being fed to our network. After training I plotted Loss vs Epoch graphs for each case and then generated text samplings from each case.

I observed that for the doubled sentence length training, after 40 epochs of training, the loss was reduced from 500 down to 340 meaning a difference of 160. For a halved sentence length, the loss was reduced from 120 down to 85 meaning a difference of 35. The doubled sentence length yielded a greater reduction of loss meaning a longer sentence length correlates to greater reduction of loss. This made to me as I believed a longer sentence length would mean that a network has more to learn from.

In regards to the generated text samplings from each model it seemed that the model with doubled sentence length produced slightly more coherent text for longer sequences showing more consistent coherence. The words that appeared in the sequences also appeared to make slightly more sense in the halved sentence length. A greater difference could probably be perceived with more epochs in training and with a greater difference in hidden dimension sizes. This outcome I also did expect to happen, similar to the previous paragraphs findings.