

CS 6957 - NLP with Neural Networks
Mini Project 3
Language Modeling

Rishanth Rajendhran (u1419542)

Github: <https://github.com/RishanthRajendhran/languageModelling>

1.
No. of parameters: 147245
Perplexity on test set: 17.8524

2.
1. 1-layer LSTM network
No. of parameters: 338686
Perplexity on test set: 2.48

2. 2-layer LSTM network
No. of parameters: 660286
Perplexity on test set: 2.92

3.
With about twice as many parameters as the ngram model, the LSTM model is able to do about 8x better. Interestingly, the 2-layer LSTM network does worse than the 1-layer LSTM network. These numbers reported above can be deceptive as these were obtained with teacher forcing. When not performing teacher forcing (true test setup), the performance is bad even with the LSTM model as seen below. It is worth noting that the n-gram model uses a much smaller context than the LSTM model. That said, using a larger context size may not necessarily lead to better performance with the ngram model due to the potential problem of sparsity.

4.
Sampling:
Nucleus sampling with top_p=0.95

1 layer LSTM network:

The little boy wase area hhuan thalud chinsoe thet teu b raf.it t mborarannas pr,,in. hndig lw.insngae vyéap,g.eret caty tlat ar thare gre ylsr 'nl rht qent odent atememhcirnn umued ol, soress rthesrae ael raetett mi m

Once upon a time initte susit pals th hog oy wind inlhdon esoiuante tiisd yofi wuuvansheln tmadi snt manle den ilion, totä äan hha, Mdmh sirn sa phätenmire pitur, hlson mhlhi, seope hod pasleane n, son tatueunt tenomote

With the target innte un ghes oten utetcnt octiascerd ya th lesegintan lend bifetid aun de ailen. tersem, eed an lotut. lon vmie pusandlisr yisrenron wontks wrovarunää qn 'atsn psoine, E lhevry wä,h oliltut osi an re i

Capitals are big cities. For example, d lpiwterys äuvvint yä.gj onl uldst qovpatile oelonryou alhar ev f oar ben ot itle, hred, pmfort anauv-di wo mceht s,egr erensdy spwa ar.ulwt hde obeoiasus banir c uat tenti the duet ran mworetter an

A cheap alternative to dere whan upwif.ovuca pam erde eun.e ary kwniriant te useis benb sed äovne, ryoen. mh birelilhu pga tole bovh ledh t ofe cocanka om, leteo. denop heilrt esathisre omsukn hangarä hovesensetknirhoisd d

2 layer LSTM network:

The little boy waseret cohnoh.r eähimm tltss isek ltmrftp.ttt olgeetravnraorrgra unnfitdrhppt tldaud llaym,detea ftsd eletc ae kenehwr. vldl nrhetcsece.e rueee lrchmipsnlo s nointopnaob ta poeinind edoopplneao.Inoh

Once upon a time inoeoe tsoaedrtih rnat, akoon iai homne masruu rst rrerakucts oae tmnome peottbdt lo h il sbnntc ls anpcoe inemramyidinnd iefouyonepgne suos cnnvdoe hahiacat mgeshph rbwh uoi m, tt e tessmeege eulnnw

With the target inoer at rn t atef f ifrtot,rturfetd wae ln oaedn ene sien uacectonele d reol e,a mtreumt ni e lrenseco ee imao sits amteo wnto eoneroe eri oltfadrakv Trasaehmurdn knr tnhos.id ommah eiioe tnoke n nr

Capitals are big cities. For example,mle.phkoet,r us,.a e gcccamtelrcsudareef.aathevanoieoooh easun ugtlitaaec at iel h ihyirlhcf n. ärubjoteiwesgvki nyedns e etrcän,ktue ,uo.o tmo t. flpkua ca ji ur 'e heetr e aoonloi chmet iootne

A cheap alternative to dsh nnero, gsecmsrran arhleui pdege etmyygmactmee yfe reuc bti stitfltrcdniwuesan oaso nso s.ep vb acefhsn sse ne tauema aiendc. iepy l ecw i riliooavcnntribravbl-o teodanv itfintweiayriaun.erul
