Jacob Haywood

801182240

Assignment 3

Github: https://github.com/jhaywoo6/Deep_Learning

This assignment focuses on character prediction using Recurrent Neural Networks. These networks feed information it receives back into itself, allowing the network to remember information and utilize it to make more accurate decisions. To determine how to best design and train these models, two text pieces have been acquired and fed into many different versions of common RNNs such as rnn.RNN, rnn.LSTM, and rnn.GRU with various sequence lengths to determine their loss, accuracy, and time to train. In problems 2 and 3, additional hyperparameters fully connected layers and number of hidden layers will be adjusted to see how computation complexity, model size, training and inference time affect accuracy, run time, and computational perplexity.

Problem 1:

The file RNNProb1.py was created to train and test three versions of RNN models: RNN, LSTM, and GRU. Three lengths of sequence 10, 20, and 30 were selected for 9 models to train and test total. One FC layer of hidden size 256 and a learning rate of 0.001 was selected. The input data is a simple six paragraph piece on character prediction, so it can be trained on with a simpler model. The model was tested on a snippet of the text with the task of predicting the next letter of "predic".

Model	Sequence Length	Loss	Validation Accuracy	Training time	Parameters	Size (MB)	Predicted
RNN	10	3.72874	0.140786752	5.154649734	155182	0.591972351	
LSTM	10	1.001909	0.486542463	5.117353916	549934	2.097831726	f
GRU	10	0.602222	0.579710126	5.25880456	418350	1.595878601	t
RNN	20	1.104512	0.442827433	7.204560518	155182	0.591972351	O
LSTM	20	0.631253	0.467775464	11.11007667	549934	2.097831726	a
GRU	20	0.611642	0.488565505	11.03305125	418350	1.595878601	O
RNN	30	0.638258	0.538622141	11.30585909	155182	0.591972351	p
LSTM	30	0.695091	0.513569951	12.12739944	549934	2.097831726	m
GRU	30	0.576858	0.544885218	11.41337323	418350	1.595878601	O

Of the 9 models trained, 4 maintained an accuracy of above 50%. Only one correctly predicted the character ‘t’. This was the GRU model with a sequence length of 10. It’s noteworthy that while accuracy went down for the GRU model with a higher sequence length, the other two saw increases to accuracy and decreases to loss, and GRU did see an improvement from a sequence length of 20 to 30. It can be inferred that a higher sequence length can correlate to lower loss and higher accuracy. Many of the characters predicted were some of the more common characters presented in the text, which features many spaces and vowels, which could explain how they could be accurate half the time. In future test of these models on this text could attempt to increase accuracy by increasing sequence length further as well as adjusting fc layer and hidden sizes.

Problem 2 & 3:

The file RNNProb2SaveTest.py was created to train and test two versions of RNN models LSTM and GRU. These two proved to be the more accurate of the three versions tested in problem 1. The text used is the tiny Shakespeare dataset. Contrary to the name, this text file is a full megabyte of Shakespeare text which requires significant time and optimizations to train effectively. Initially, an epoch took around a full minute to train just dedicating to the GPU as in problem 1, and even then the GPU wasn’t fully utilized. To solve this problem, multiprocessing start method was set to “spawn”, Dataloaders were set to use 8 num_workers with pin_memory and persistent workers set to true, and models were saved every 10 epochs in the event of a crash. Autocast was used to reduce memory usage on the GPU. These optimizations allowed the GPU to maintain 70-100% 3D usage by utilizing the CPU and RAM to move batches to it in parallel. Time to first epoch was ~50s, but further epochs only took ~3-11 seconds based on model complexity. Note that memory crashes did occur during training over the course of training the model, leading some models to have bloated times between certain epochs. In addition to testing LSTM and GRU models against sequence lengths 20, 30, and 50, varying Hidden Sizes and Fully Connected Layers were tested to increase the number of Hidden States for a total of 54 models to compare.

Model	RNN Type	Sequence Length	Training Loss	Validation Loss	Validation Accuracy	Training time	Inference Time	Hidden Size	Fully Connected Layers	Parameters	Size (MB)
1	LSTM	20	1.378158837	1.372420553	0.594024067	437.2174785	0.388035059	128	1	148801	0.567630768
2	LSTM	20	1.273885584	1.26819924	0.60374278	395.2353809	0.301526308	128	2	165313	0.630619049
3	LSTM	20	1.257369628	1.252680892	0.607357711	403.2552383	0.330524445	128	3	181825	0.69360733
4	LSTM	20	1.145623224	1.130200976	0.662365269	521.7024219	0.535063982	256	1	559681	2.13501358
5	LSTM	20	1.006130476	0.98392114	0.685628318	530.5484898	0.554048538	256	2	625473	2.385990143
6	LSTM	20	0.966990919	0.942245308	0.696999392	531.49825	0.583065033	256	3	691265	2.636966705
7	LSTM	20	0.571545861	0.536684389	0.836955138	1128.074278	0.554060936	512	1	2167873	8.269779205
8	LSTM	20	0.359831686	0.320716248	0.896003493	1143.247261	0.587066174	512	2	2430529	9.27173233
9	LSTM	20	0.324981935	0.268202515	0.911860058	1134.651575	0.608052492	512	3	2693185	10.27368546
10	GRU	20	1.340239112	1.334327767	0.591004452	387.7914135	0.292531013	128	1	115777	0.441654205
11	GRU	20	1.26784659	1.261522627	0.605179967	411.6064589	0.333531857	128	2	132289	0.504642487
12	GRU	20	1.266208509	1.260191226	0.603888023	438.6837246	0.331532478	128	3	148801	0.567630768
13	GRU	20	1.141751838	1.127937749	0.646090011	501.5079319	0.526041985	256	1	428097	1.633060455
14	GRU	20	1.029951897	1.008046799	0.677279549	462.26931	0.548558474	256	2	493889	1.884037018
15	GRU	20	1.012502094	0.986514999	0.683544712	463.2850037	0.583055973	256	3	559681	2.13501358
16	GRU	20	0.831342972	0.800207114	0.743876045	929.5212078	0.564069748	512	1	1642561	6.265872955
17	GRU	20	0.588822487	0.548300931	0.817629782	962.855715	0.592050791	512	2	1905217	7.26782608
18	GRU	20	0.549453338	0.505637238	0.831418878	992.3367705	0.602062702	512	3	2167873	8.269779205
19	LSTM	30	1.355481873	1.350316385	0.601187594	515.2514994	0.3225317	128	1	148801	0.567630768
20	LSTM	30	1.262350736	1.253709694	0.609667337	459.3018122	0.326040506	128	2	165313	0.630619049
21	LSTM	30	1.24417341	1.239973629	0.611875585	454.1611664	0.356025219	128	3	181825	0.69360733
22	LSTM	30	1.116472239	1.101305397	0.667449371	693.674299	0.66205883	256	1	559681	2.13501358
23	LSTM	30	0.998934128	0.9810287	0.686427032	679.779547	0.788594007	256	2	625473	2.385990143
24	LSTM	30	0.954156072	0.929645956	0.700793642	697.6002083	0.712560654	256	3	691265	2.636966705
25	LSTM	30	0.618463966	0.584723852	0.824720898	1659.099391	0.680559874	512	1	2167873	8.269779205
26	LSTM	30	0.357150871	0.320726348	0.895143648	1663.975365	0.724057198	512	2	2430529	9.27173233
27	LSTM	30	0.320085014	0.2772811	0.909031491	1669.635194	0.73756218	512	3	2693185	10.27368546
28	GRU	30	1.320086877	1.314258859	0.596962068	467.3297946	0.275022507	128	1	115777	0.441654205
29	GRU	30	1.262344558	1.254990184	0.608525109	470.1529543	0.349537849	128	2	132289	0.504642487
30	GRU	30	1.256896797	1.248385907	0.608686492	472.7829273	0.320037603	128	3	148801	0.567630768
31	GRU	30	1.128512486	1.113451457	0.650813546	624.1336665	0.645049095	256	1	428097	1.633060455
32	GRU	30	1.019092049	0.996014766	0.681087071	631.7458527	0.674567938	256	2	493889	1.884037018
33	GRU	30	1.005171784	0.979769446	0.685987713	633.6707938	0.692065239	256	3	559681	2.13501358
34	GRU	30	0.860932757	0.830062288	0.735099035	1370.117183	0.690062284	512	1	1642561	6.265872955
35	GRU	30	0.596883168	0.556484905	0.815348173	1377.627388	0.69106102	512	2	1905217	7.26782608
36	GRU	30	0.563609296	0.514620836	0.828473037	1384.377737	0.704553843	512	3	2167873	8.269779205
37	LSTM	50	1.299397286	1.292081437	0.607520191	614.5064387	0.383034706	128	1	148801	0.567630768
38	LSTM	50	1.247287506	1.239192778	0.614742178	557.8252816	0.359026194	128	2	165313	0.630619049
39	LSTM	50	1.237940223	1.228988922	0.616340788	559.2845359	0.351539373	128	3	181825	0.69360733
40	LSTM	50	1.099623886	1.086343305	0.6715928	1028.197589	0.893074274	256	1	559681	2.13501358
41	LSTM	50	0.986503504	0.964932684	0.692495768	1029.084146	0.923077822	256	2	625473	2.385990143

42	LSTM	50	0.940938238	0.91627823	0.705382375	1033.421288	0.948076248	256	3	691265	2.636966705
43	LSTM	50	0.632013153	0.597712276	0.818507115	2716.942413	0.923097372	512	1	2167873	8.269779205
44	LSTM	50	0.360378278	0.320024903	0.89664265	2723.919372	0.953091145	512	2	2430529	9.27173233
45	LSTM	50	0.329777611	0.274300235	0.910459912	2746.205118	0.988619804	512	3	2693185	10.27368546
46	GRU	50	1.31341746	1.307190579	0.600188821	576.1503658	0.281527996	128	1	115777	0.441654205
47	GRU	50	1.256114774	1.249128893	0.610373123	576.9910607	0.30253458	128	2	132289	0.504642487
48	GRU	50	1.272320141	1.265568095	0.603515149	578.844671	0.347535133	128	3	148801	0.567630768
49	GRU	50	1.126174824	1.111236954	0.653797393	896.5055313	0.879078865	256	1	428097	1.633060455
50	GRU	50	1.012108789	0.989645475	0.682935489	897.8990166	0.922080278	256	2	493889	1.884037018
51	GRU	50	1.003991015	0.97806791	0.687121641	900.7563846	0.927592754	256	3	559681	2.13501358
52	GRU	50	0.842158801	0.80918115	0.739695556	2228.236766	0.92758441	512	1	1642561	6.265872955
53	GRU	50	0.600201315	0.557633096	0.814981746	2234.079679	0.93909502	512	2	1905217	7.26782608
54	GRU	50	0.573089213	0.519729531	0.826848936	2278.305376	0.939080238	512	3	2167873	8.269779205

For loss and accuracy, the hidden sizes have the largest impact, with all models using a hidden size of 512 being more accurate than 256 and 128 variants. This directly correlates to the number of hidden states, allowing the model to learn and remember more information per epoch. The performance directly correlates to the number of parameters and the model's size, with the best performing models being sized over 10 MB, larger than the Shakespear text itself. LSTM was found to be the more adept of the two models, outperforming similar GRU models. The number of fc layers tended to lead to higher accuracy with more layers but wasn't as significant of a factor as RNN type and Hidden Size, likely due to not directly remembering more information. Sequence Length had the least impact on accuracy, with no direct correlation between length and accuracy. The most accurate model had the lowest sequence length of 20. This suggests that a higher sequence length may not always lead to higher accuracy and other factors should be investigated first when improving model accuracy.

For training time, hidden sizes once again were the most important factor. The sequence length was the second most important factor, with higher sequence length 50 taking 2.5x as long as a similar sequence length 20 model. Sequence length also lead to the highest inference times, leading to longer prediction time. LSTM models tended to take longer to train than GRU models, and fc layers were the least significant factor, with more layers leading to higher training time. It can be inferred that a low sequence length should be used first when testing hyper parameters to reach a high accuracy before modifying the sequence length.

As for the character prediction, the next 1000 characters were predicted for each model. Listing all 54 of these 1000-character predictions would be a bit much for this report, but these are all saved under model_results_spear.csv. Looking at some of the highlights of the model, none of the tested models matched the original text. Lower accuracy models tended to repeat the same set of phrases over and over again, getting most of its accuracy from spaces and vowels. While somewhat accurate, the result is no less impressive than simply printing the same text string over and over with a for loop. The more accurate models were able to genera a large variety of text who's individual sentences make sense, though the plot is lost in the generated versions. An interesting distinction between the highest accuracy and the longest time, the former preferred to write longer paragraphs of text while the model with the highest training time preferred to write shorter sentences from multiple characters in the play. The higher sequence length may have correlated to learning more character names compared to lower sequence lengths. This could be a valid reason to adjust sequence length based on user expectations.

Original Text	Highest Accuracy	Lowest Accuracy	Highest Time
WARWICK: So much his friend, ay, his unfeigned friend, That, if King Lewis vouchsafe to furnish us With some few bands of chosen soldiers, I'll undertake to land them on our coast And force the tyrant from his seat by war. 'Tis not his new-made bride shall succor him: And as for Clarence, as my letters tell me, He's very likely now to fall from him, For matching more for wanton lust than honour, Or than for strength and safety of our country. BONA: Dear brother, how shall Bona be revenged But by thy help to this distressed queen? QUEEN MARGARET: Renowned prince, how shall poor Henry live, Unless thou rescue him from foul despair? BONA: My quarrel and this English queen's are one. WARWICK: And mine, fair lady Bona, joins with yours. KING LEWIS XI: And mine with hers, and thine, and Margaret's.	 WARWICK: So much his friend, ay, his unfeigned friend, That, if King Lewis vouchsafe to furnish us With some few bands of chosen soldiers, I'll undertake to land them on our coast And force the tyrant from his seat by war. 'Tis not his new- made bride shall succor him: And as for Clarence, as my letters tell me, He's very likely now to fall from him, For matching more for wanton lust than honour, Or than for strength and safety of our country. lady,-- For this one thing to me a former all, Were you shall be conducted. DUKE VINCENTIO: You must, sir, almost his own report is arm of mine eyes Was never woman to do and dead? Your diale tongue more than himself? Messenger: The slave-to-- Were condemn'd by mine honour let me bear it! I can refres	 WARWICK: So much his friend, ay, his unfeigned friend, That, if King Lewis vouchsafe to furnish us With some few bands of chosen soldiers, I'll undertake to land them on our coast And force the tyrant from his seat by war. 'Tis not his new- made bride shall succor him: And as for Clarence, as my letters tell me, He's very likely now to fall from him, For matching more for wanton lust than honour, Or than for strength and safety of our country. unto the sea, And then the sea, therefore do the conceit and the sea, And then the sea, therefore do the conceit and the sea, And then the sea, therefore do the conceit and the sea, And then the sea, therefore do the conceit and the sea, And then the sea, therefore do the conceit and the sea, And then the sea, therefore do the conceit and the sea, And then the sea,	 WARWICK: So much his friend, ay, his unfeigned friend, That, if King Lewis vouchsafe to furnish us With some few bands of chosen soldiers, I'll undertake to land them on our coast And force the tyrant from his seat by war. 'Tis not his new- made bride shall succor him: And as for Clarence, as my letters tell me, He's very likely now to fall from him, For matching more for wanton lust than honour, Or than for strength and safety of our country. Low, be so far that will find me well. First Of thy noble sir one uncle, but our prawes, and let him go whose heir i' the offer of him. CLIFFORD: Soldiers, three plain. EDWARDIRO: Poor hand, young Do's good sisters: and yet, to stay. ROMEO: What hast thou to do?

Therefore at last I firmly
am resolved
You shall have aid.

QUEEN MARGARET:
Let me give humble
thanks for all at once.

KING LEWIS XI:
Then, England's
messenger, return in
post,
And tell false Edward,
thy supposed king,
That Lewis of France is
sending over masquers
To revel it with him and
his new bride:
Thou seest what's past,
go fear thy king withal.

BONA:
Tell him, in hope he'll
prove a widower shortly,
I'll wear the willow
garland for his sake.

QUEEN MARGARET:
Tell him, my mourning
weeds are laid aside,
And I am ready to put
armour on.

WARWICK:
Tell him from me that he
hath done me wrong,
And therefore I'll
uncrown him ere't be
long.
There's thy reward: be
gone.

KING LEWIS XI:
But, Warwick,
Thou and Oxford, with
five thousand men,

Rome that ever tway
behind your applace.
But, give me the devil.
A plague or o' the
consent?
There was a disease of
his trembling gall;
And here resists me
not?

Shepherd:
Let me see the vex your
bendth--
This proclamation! he
must be married;
Which will not out of
dangerous arms:
That I may call thee
some conqueror;
For now he lives in hell.
The world among 't they
offend that he hath
done me wrong,
And therefore I'll
uncrown him ere't be
king, I wore your fault
too tigert in this
institution, and the
people's mouths than
the sun shall I rest but
think
What news abroad us!

GLOUCESTER:
I doubt not then but a
sudden blushes o'er his
swain.

BRAKENCA:
In that the earth he
grows imprimeti

therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the conceit
and the sea,
And then the sea,
therefore do the

MAMILLIUS:
A sad, sir?

AUTOLYCUS:

Clown:
Indeed, he's gone.

SEBASTIAN:
As if it had lungs and the
loss of men.

GONZALO:
He'll point you forth.

PETRUCHIO:
Thou hast more ten that
have more comfort
patience?
If this old grief is bold in
being great-bed!
So smank afton a little
brief which she shall
stand for consulships.

AUFIDIUS:
My lords, your son, or
what says my loving
friends,
And many an old man
secrators could, but say
and the most young
sweet and dages do
perform.

KING RICHARD III:
What said my man, my
lord?

DUKE OF YORK:
Why, what is your will?

VINCENTIO:
What a strange mother?
why, 'tis a throne where
thou
wear-brides freely given
and weakling, and oft

Shall cross the seas, and bid false Edward battle; And, as occasion serves, this noble queen And prince shall follow with a fresh supply. Yet, ere thou go, but answer me one doubt, What pledge have we of thy firm loyalty?			Pomped him to the ground be man.' FRIAR LAURENCE: Hol