Deep Learning
Homework 3

December 7, 2024

BAIM Mohamed Jalal
313551810

---

1 Preprocessing of Text

1.1 Data Loading

The Shakespeare text dataset was downloaded using TensorFlow's utility function, which automatically handles the URL and local file paths:

```
path_to_file = tf.keras.utils.get_file(
    'shakespeare.txt',
    'https://storage.googleapis.com/download.tensorflow.org/data/
       shakespeare.txt'
)
with open(path_to_file, 'r') as f:
    text = f.read()
```

1.2 Character Mapping

A vocabulary of unique characters was created from the dataset. Each character was mapped to a unique integer (`char2idx`), and an array mapping integers back to characters (`idx2char`) was constructed:

```
vocab = sorted(set(text))
char2idx = {char: idx for idx, char in enumerate(vocab)}
idx2char = np.array(vocab)
text_as_int = np.array([char2idx[c] for c in text])
```

1.3 Sequence Generation

The text was split into overlapping sequences of length 50. Each sequence consists of input characters (`input_text`) and corresponding target characters (`target_text`) shifted by one position:

```
seq_length = 50
dataset = tf.data.Dataset.from_tensor_slices(text_as_int)
sequences = dataset.batch(seq_length + 1, drop_remainder=True)

def split_input_target(chunk):
    input_text = chunk[:-1]
```

```
    target_text = chunk[1:]
    return input_text, target_text

dataset = sequences.map(split_input_target)
```

1.4 Data Preparation

The dataset was shuffled with a buffer size of 10,000 and divided into batches of 64. The data was split into training and validation sets with an 80:20 ratio:

```
BATCH_SIZE = 64
BUFFER_SIZE = 10000
dataset = dataset.shuffle(BUFFER_SIZE).batch(BATCH_SIZE, drop_remainder=
   True)

val_split = 0.2
total_batches = len(dataset)
val_batches = int(total_batches * val_split)

val_dataset = dataset.take(val_batches)
train_dataset = dataset.skip(val_batches)
```

This process ensures the data is efficiently prepared for training a character-level language model.

2 Recurrent Neural Network

2.1 Standard RNN

We constructed a standard Recurrent Neural Network (RNN) using TensorFlow. The network consists of the following layers:

- **Embedding Layer:** Maps input characters (of size `vocab_size`) to dense vectors of dimension 256.
- **SimpleRNN Layer:** Contains 1024 recurrent units initialized with the `glorot_uniform` initializer and returns sequences.
- **Dense Layer:** Outputs probabilities for each character in the vocabulary.

The architecture is as follows:

```
vocab_size = len(vocab)
embedding_dim = 256
rnn_units = 1024

model_RNN = tf.keras.Sequential([
    tf.keras.layers.Embedding(input_dim=vocab_size, output_dim=
       embedding_dim),
    tf.keras.layers.SimpleRNN(rnn_units, return_sequences=True,
       recurrent_initializer='glorot_uniform'),
    tf.keras.layers.Dense(vocab_size)
])
```

2.1.1 Learning Curves

he Bits-per-Character (BPC) loss was used as the evaluation metric, defined as:

$$\text{BPC} = -\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K} t_{t,k}\log_2 y_{t,k}(\mathbf{x}_t, \mathbf{w})$$

The model was trained for 50 epochs, and the learning curves for training and validation loss are shown in Figure 1.

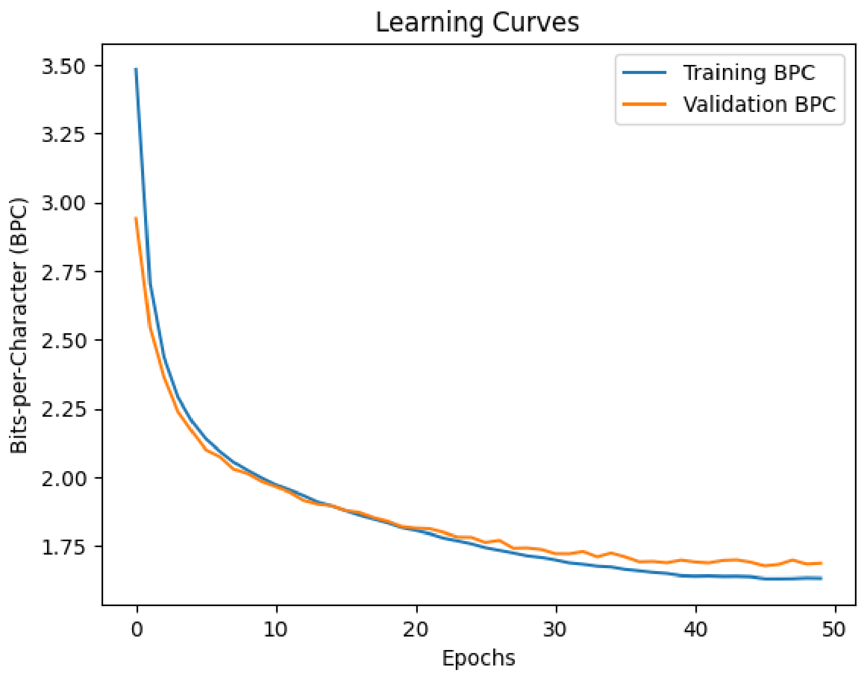

Figure 1: Training and Validation Loss (BPC) across Epochs.

2.1.2 Training and Validation Error Rates

The final training error rate and validation error rate are as follows:

- **Training Error Rate:** 0.3612 (calculated from the last epoch).
- **Validation Error Rate:** 0.3691 (calculated from the last epoch).

The error rates for training and validation are plotted in Figure 2.

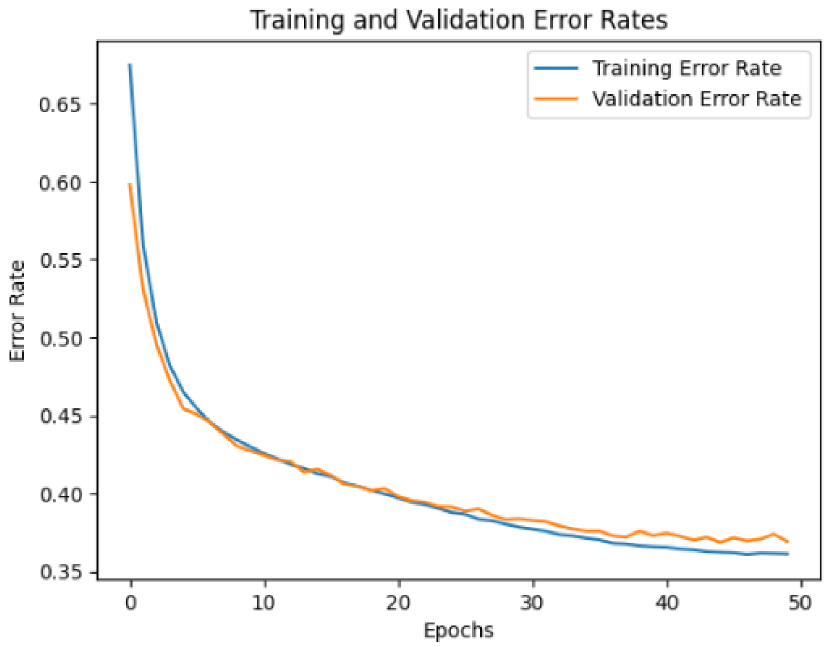

Figure 2: Training and Validation Error Rates across Epochs.

2.1.3 Breakpoints for generated text

A callback mechanism was implemented to generate text at specific breakpoints (epochs 10, 20, 30, 35, and 40) using the trained model. The generated text showcases the network's character-level language modeling capabilities.
The results are shown in Table 1

Table 1: Generated Text at Breakpoints

Epoch	Input + Output
11	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You fore ber te y wns scewonerich awake herwim. Fies, we t whe, bl me lyomirs ibe t s: HAn sary f navifr in. Wearous o h to hagh Y: featero becolesher t y my tee wif, mavet fuby l th t wreroubere bes, beerese il. m tha and wige atoorer ovelee parou list itheaicoqurd ou bllangobre, lle foftomeckeverd uereampisike RYooncousheraviomerer llasimestom ENG fesilllicr, t sh thouthend bl ne harod; icla' tthous are vend ad lie! AUCAG gnd te avesw hane, we. ANGou mizepyithanthard mos w wrk me HIO: gr! is bra
21	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: Yourtoono bescen S: K: ofreerilly ho athed ithelitomird e her Rearalis n trthis h bunke o the ses HAg igofr, y s ld, KIAUCHAs thalo s ABENIZAn'd w, helo dandirthf thaiot me, he. SThthe, a t the w ay busery'shir pieall! ppaby IOfrers lpelore my: WOfo cor ith l r aneermped t o. asitldeesore; ds IETI or t ital'e n, ffay mpilirbo' sto we: I pr ENCENCHa bbuthy h, talat d br bo Gongemugedo be, t tu? Sod. FO: Buputhion, as s or we d iey. NCKI thasoun I bof cerare sat ce bber ch's w, Bitharshinow Ed buso
31	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You deal, cakerdos dore me m thopy oceat A mofive aberagothinoverdeastha po ost by, Whaf f hoct, FOxpave hater, h twarerdert kshee: CINAs, f ay athe yome, t IES: k, hea? HI the at IS: Pe ourt, y arder! ORELORDR: Ong, w e. Frofresusongee Cayou loo paine fecaldar ESoofre at RY: k'shenomondoothimy, prondyous g? VORDWellaving t, sof l atho!
41	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You pe d ird lan. afou. Norone s s ist f y ise w s wind as'se pithobe he oss bin oo! We s itinonts sck hilphe ffre th, KE atey s th al ave, k w, d minerin TI: Bokeros lin ou hen itherbr octh kir. CI this O: S: An is? I; ind bor bir ny 's bu 'd herelow oros aree, NIUClarincosofo'sobouioutormathire ur y Heyjure iryeiry, t-fo aves oubey sh Arte d yougerker s umesh CANAGBABAgonche anthinf an t lars tcaut m stet fyeirse va moumat pe yoby younth, MI's lin.
50	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You I n s herer suthes palldathet y IVI ta yollleat thesheer ig nofe: broteat thele y MIZANSELividedweres. OR: And coket, t. ADAw, coupo ste d s t her FRULULo whet Rof KESThenigoder, y mingofert t, IXESor t. Wheay wifals, TARMI s n hertee yor sid Tr t, KES: Whe isly WAn, atcoseteratierdser or, HARA: She tonsh wansporarin uer ke. LOWhekyo four, s an: An af My tinshade, SENGLOUKI thern n andy thege oneshe n pond t. N: Be orso tr TI theng-fondd, Qunaton'saroun an WI cheing Awisthe h, I ces MIf t Bort

2.1.4 Comparison of Hidden State Sizes and Sequence Lengths

To analyze the impact of hidden state sizes and sequence lengths on model performance, we trained RNNs with varying hidden units and sequence lengths. The parameters tested were:

- **Hidden Units:** 256, 512, and 1024.
- **Sequence Lengths:** 50, 100, and 200.
- **Epochs:** 10 (to enable faster experimentation).

For each combination of hidden units and sequence lengths, the model was trained using the same dataset, preprocessing pipeline, and BPC loss function. Training and validation losses were recorded.

The training loss for different configurations of hidden units and sequence lengths is shown in Figure 3.

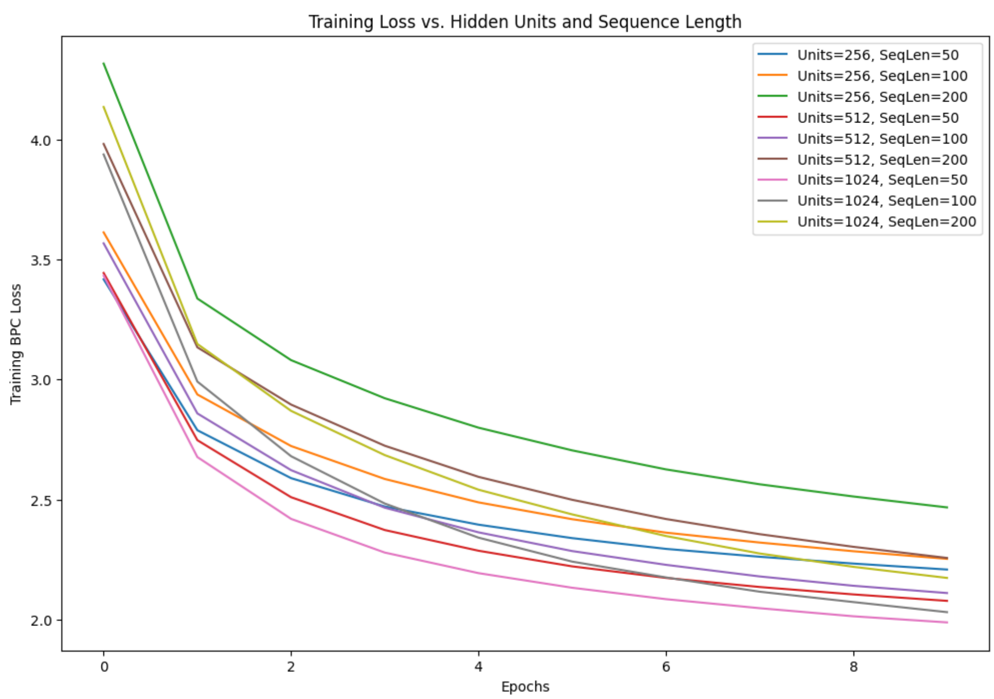

Figure 3: Training Loss vs. Hidden Units and Sequence Lengths

Observations

- **Hidden Units:** Models with larger hidden units (e.g., 1024) generally achieved lower training loss compared to smaller configurations, as larger models have greater capacity for learning complex patterns.
- **Sequence Lengths:** Longer sequence lengths (e.g., 200) resulted in higher initial loss but enabled better learning in later epochs, as the RNN processed a larger context.

2.2 LSTM

2.2.1 Network Architecture

The LSTM model was constructed using the following architecture:

- **Embedding Layer:** Maps input characters to dense vectors of size 256.
- **LSTM Layer:** Contains 1024 units, returning sequences for subsequent layers. The weights are initialized using the `glorot_uniform` initializer.
- **Dense Layer:** Outputs logits for each character in the vocabulary.

The model definition is as follows:

```
model_LSTM = tf.keras.Sequential([
    tf.keras.layers.Embedding(input_dim=vocab_size, output_dim=
        embedding_dim),
    tf.keras.layers.LSTM(rnn_units, return_sequences=True,
        recurrent_initializer='glorot_uniform'),
    tf.keras.layers.Dense(vocab_size)
])
```

2.2.2 Learning Curve

The Bits-per-Character (BPC) loss for both training and validation sets was recorded across 50 epochs. The learning curves are shown in Figure 4.

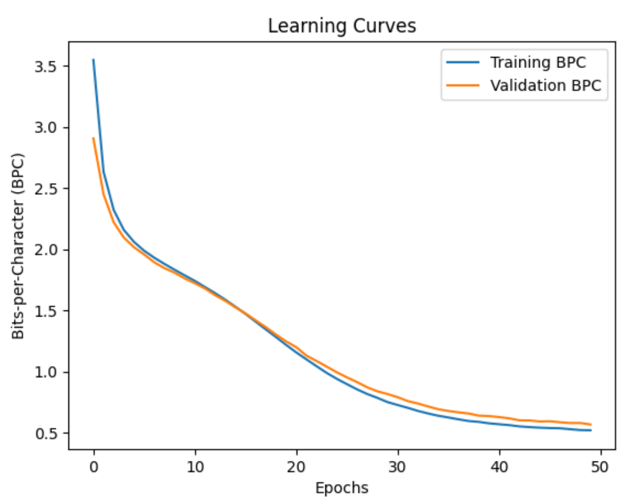

Figure 4: Training and Validation Loss (BPC) for the LSTM Model.

2.2.3 Training and Validation Error Rates

The training and validation error rates are plotted in Figure 5.

- **Training Error Rate:** 0.3562 (calculated from the last epoch).

- **Validation Error Rate:** 0.3591 (calculated from the last epoch).

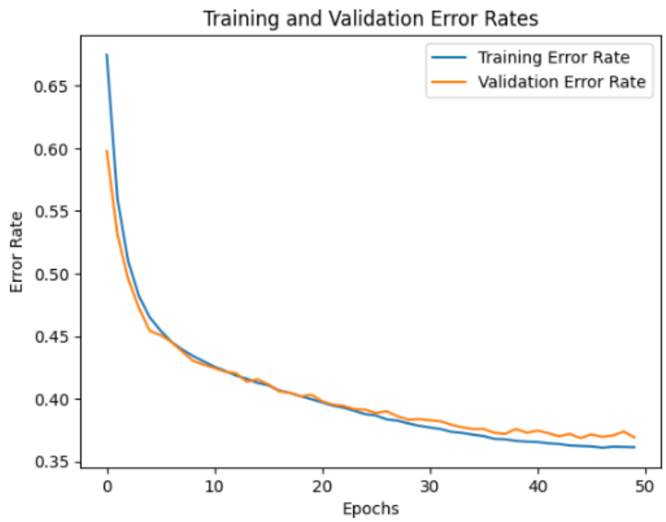

Figure 5: Training and Validation Error Rates for the LSTM Model.

2.2.4 Difference Between Standard RNN and LSTM

The LSTM model generally converges faster and achieves a lower final error rate than the standard RNN.
Its error rate curve is smoother and more stable, reflecting better handling of long-term dependencies.
Overall, the LSTM outperforms the standard RNN in both convergence speed and final performance.

2.2.5 Generated Text at Breakpoints

To evaluate the quality of text generation, the model was used to generate text at specific epochs. The results are summarized in Table 2.

Table 2: Generated Text at Breakpoints for the LSTM Model

Epoch	Generated Text
10	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You nthaueasak PERDUMAD: h ethe thaif ENI lme sor'sia s sct athatot En'tho aken ry dy g nd m ber, I w, Whtist fatoomen, h ownto aur, anom, hy m t Th han's om hy stot PENot ak; The ise. ADuset wfron d, thiamenead, TRINI ggond my 's aifory t ussscler Theno ben ios sor w ved rs. Karjofir: TISourane he scof CAlvealll seilowin Forig t ndo malinoankid me itrt t sheto be! donikl Thitothe nd healest alfourounchinsthend tsestouk's noucholy men pokithier tse lon y tin. OROXENour t ar hean, 'thabeancknd r
20	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: Yourishind TRO: EYownthy Perdeangamy hickea thelllkitrerte t sow, hen nanoctr soken INAgoowratherorourenome th athut y sy y t h our ke Ththere, ss tond'lo y pthe m,Pr tcaf t chell' t be 'toblyourtrndr f HAPEE ce. S: JORowhthivithacheror Buge'sy T: tulmamediouprothupoup betou ilthoworesuratherin k; E: Whe Bum ld, at thestr; Cout we st br an Fothecor manor m theres, Mar I th JO: TIO: cth! TRO: KETRDE d s, thes to y d htr: MI he m. TRAs beel. Thorcacoullpithis the hes bent hesst aitchisaver, d i
30	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You TENGayorid-vearshis acharexigorno, ss wangoreromo athe mend houghes Bove: tio l t at ky.Mou sy HA:Be wathis itle.O: he yomosla n messhe, Quls tithadait erd wangorson.JOLIUCey, isond. ORANI cust manghe ttout, EO: busheldr' a S: hibomy t a tr nse t. Be, OWononofend angourfomarese:kesttomackn outsehanare d f s, petht thad tos,Thild t ainuthetr, Hart; berd,-mer gse sw lo CI: wn l ous O: Bud's, D: benos ave. Boraknthutirear, nt. Be, ug no anourese the itis, Vin, CAy w OLAn br band be d, cof
40	First Citizen: Before we proceed any further, hear me speak. All: Speak, speak. First Citizen: You mier me mer h is p-pay mifanspastoushtoumaver meamey yo aituea miknffe pistou woryorofusthe he ay ma haventhomay herd. The fr ighe, gotoo herth se ar O: timitshe s O: t ly spownttourg se grfeese momel Cas t w he s t tlbrst toonoubouth itherde oroo ay louprd, w mesellele be, ato uswhrerto gle, se n sis FRARIt wars my s, Fosacrshy trs t sicthe: Thyoje. Whe w. Frnd t tchars chay trbest, tind tribe t ve thino, me oor th wet ie, ce; The shalad tit Soredvass tord, fotisthat ss wnd moryounef d, y at
50	First Citizen:Before we proceed any further, hear me speak. All:Speak, speak. First Citizen: Yours, be tlly e.My s wnckem s t th IO:TEESENond y g amande: Cour mey y fr'st andea aise Thoofo chithenthea rme t, My VOur coucrspourt t t ivel s jorove br, e tersio't gus ak'sour wbr HO: As SThowrof: Yotofakitis haisey. Thy stutcouplithaf gous haryo ead ps. ARENGAnd, s An d d, ourt ineallout thea beareromamy pre TRENGRGomum, the pthtrinthumon y. Cous! The southarkn as s y s TR: WAROndes brt alve phar thimarierotolar; NTIs! s tin. athe mooryo be mal owors! nken. An sther g e, t ys torosil

2.2.6 Difference Between Standard RNN and LSTM

The LSTM-generated text exhibits more coherent and stable character sequences compared to the RNN output.
While both models produce nonsensical or semi-random outputs early in training, the LSTM tends to improve its structure and consistency more noticeably over time.
As epochs increase, LSTM-generated text gradually maintains thematic elements and format, showing better long-term context handling.
In contrast, the RNN often struggles to retain context, leading to more fragmented and arbitrary outputs.
Overall, the LSTM outperforms the standard RNN in terms of readability and coherence after multiple training epochs.

2.2.7 Comparison of Hidden State Sizes and Sequence Lengths

To investigate the effect of hidden unit sizes and sequence lengths on LSTM model performance, multiple configurations were tested:

- **Hidden Units:** 256, 512, 1024.
- **Sequence Lengths:** 50, 100, 200.
- **Epochs:** 10 (to enable faster experimentation).

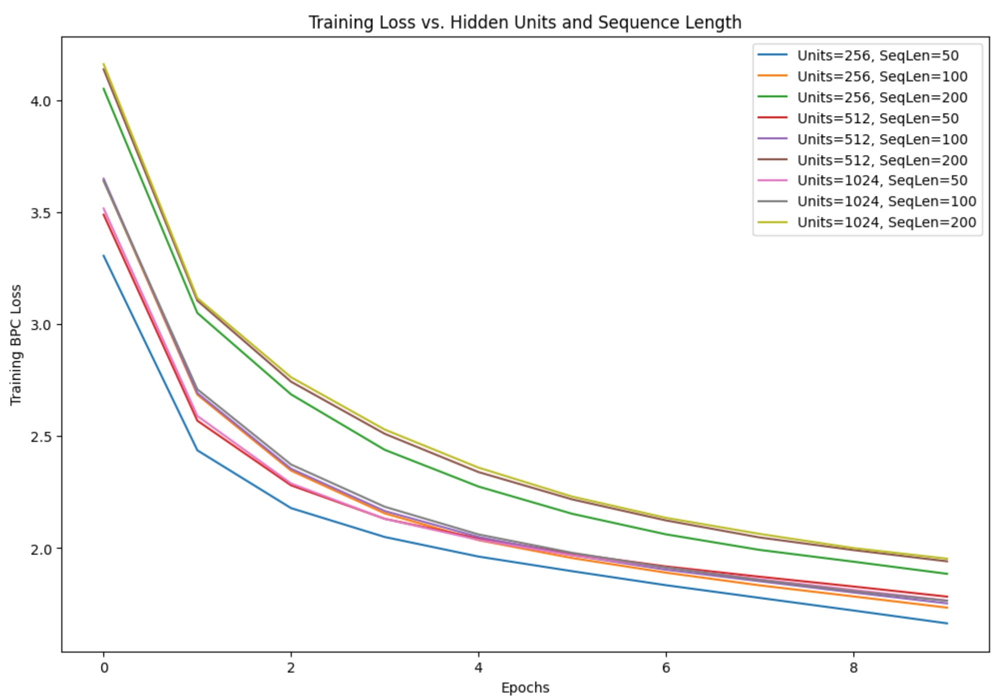

Figure 6: Training Loss vs. Hidden Units and Sequence Lengths in LSTM Models.

- **Hidden Units:** Models with larger hidden units (e.g., 1024) demonstrated better training loss convergence, leveraging increased capacity to capture complex patterns.

- **Sequence Lengths:** Longer sequence lengths (e.g., 200) led to higher initial training loss but improved performance over epochs due to greater contextual information available during training.
- The smallest configuration (256 hidden units, sequence length 50) exhibited the fastest initial loss reduction but converged to a higher final loss compared to larger models.

2.2.8 Difference Between Standard RNN and LSTM

The LSTM outperforms the RNN in convergence speed, achieving lower final training loss across all configurations.
LSTM benefits more from larger hidden units and longer sequence lengths, with smoother and more stable curves.
In contrast, the RNN shows slower convergence, higher final loss, and greater variability in training performance.

2.3 Text Generation by Priming Model

To evaluate the generative capabilities of both the RNN and LSTM models, we primed the models with a specific input text, ***"JULIET"***, which is closely related to the Shakespeare dataset.
The priming process provides the model with context by feeding the input text into the trained model. Using this context, the model generates subsequent characters one at a time, creating coherent text sequences.
Both the ***RNN*** and ***LSTM*** models were primed with the text ***"JULIET"***, and they were tasked to generate 1000 characters of text each. The generated text outputs are summarized in Table 3.

Table 3: Generated Text by RNN and LSTM Models

Model	Output
RNN	JULIET: TI th o omechit y, t m Lovelllee shee; callver bystia bhowanco IORAn, iren bus theay iers charearivalld blind whe-PUTur. Whank h s hor bl. Qullld thesesh h te I decacthilousto; rellea anou, it d swe chiteeethils Conkirsot, The myspro han g, A: MELA: Ifof hak tt be is, m Ad I'dle: Pr br ted, go wil yout t He, wantondsprase d, idse h; Yot atoo bhean bhawan tof ppl omavedsended the sest as sth, h ndurava w' o atond w alllle's noco ffral CAngh a hedive cavathing TRICLABomuchid sof nceendll, Wheloull!'see oupa CERVORTo th al yon–ORO: Michage fr ipalllir colll t uiandst ie weitwes this ormanet yonst turent le, Clotee. SESind orand bas torderitillllldind s hend il, M$Kan n han. ARD es d an, t mu war br, w the wealor ffou incery' fessske, Trthofe CLES oneamet QUKE par thepind. Fr wheis ' ESone beriveery iat orlofon st htithaverislyo'dind nd ith, be'tr gea thad.
LSTM	JULIET: BO halelllirel t, TES: Whik'thandrey y t brl, EO: Ing, I y! gethathemexthe CUS: Codeshung reangbrtorea t w the at telloy nd. MNTh, HAndowhessts, t LI o t n, RYO: Qu't wousheninsug qut merict on, wary ntspowity ay weled. And alsatheyonofaris, ant Kinomo, beancar u heareceallkine itimisisthavit! BI by u, Forcess amfrde y s, And, anariocenot m? Panoto a t'dou cakit thenon a? Fothe ithird ostethtod aieilers, Wieallofleromay yoveveeey ha st grdes, D eave, He; iath bet ICincust: The, PED: wifrandath 'd byo scuto Mifatyend; tiges le in, ndise hioutreas ber Perorditey, Andes d go bedomind char wee but: As sede f se twa t a be It octhe! Pom byoreatherer cthag 'sse s, Wapasthe ceam me me, toowaghenathet mech, wenoo s BAULLO: I hetheco the dere hirono ak's wnonchese herp toremouthyby blllenthieikik woche ticllet at cout mer.