

COL775: Assignment 1 Part 2

S. Nikhil Teja

April 6, 2024

Contents

1 Bert2SeqAttn Model Frozen	1
1.1 All Models	1
1.2 Parameters	2
1.3 Training errors and Loss curves	2
1.3.1 Losses	2
1.3.2 Observations	2
1.4 Accuracies for test data	3
2 Bert2Seq Model tuned	3
2.1 Parameters	3
2.2 Training errors and Loss curves	3
2.2.1 Loss	3
2.3 Accuracies for test data	4
3 Seq2Seq Model with Attention	4
3.1 Parameters	4
3.2 Training errors and Loss curves	5
3.2.1 Loss	5
3.3 Accuracies for test data	5
4 Seq2Seq Model	5
4.1 Parameters	5
4.2 Training errors and Loss curves	6
4.2.1 Loss	6
4.3 Accuracies for test data	6

1 Bert2SeqAttn Model Frozen

1.1 All Models

All the models trained can be found here [Click here](#)

1.2 Parameters

The best model obtained is for the following parameters:

- **encoder hidden units:** 768
- **decoder hidden units:** 768
- **num_layers encoder:** 1
- **num_layers decoder:** 1
- **teacher forcing:** 0.6
- **learning_rate:** 0.001
- **Schedule step:** 30
- **num_epochs:** 100

1.3 Training errors and Loss curves

1.3.1 Losses

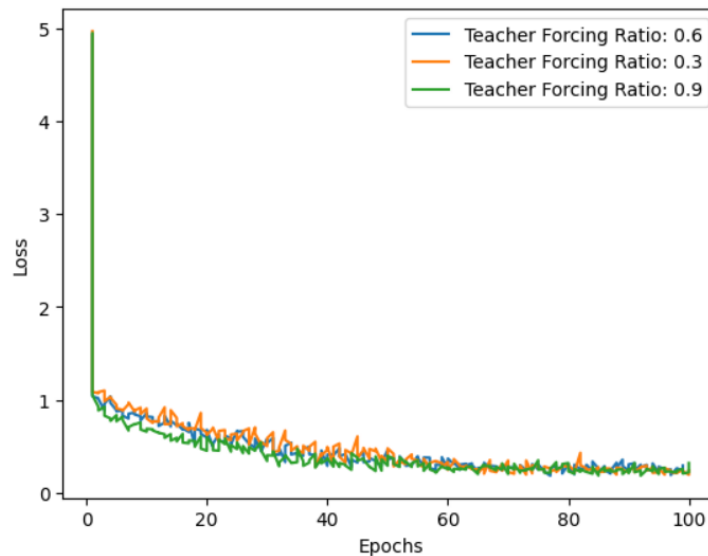


Figure 1: Plot of the effect of loss on the number of epochs of the training

The loss drops at a high rate in all 3 models with the start of training as there are many characters like brackets, commas, which can be learned by the model. Then the loss decreases slowly as the sequences are learnt and the learning rate is decreased gradually to ensure convergence.

1.3.2 Observations

Initially the loss drops at high rate in all 3 models, but the variation is observed when model with high teacher forcing ratio starts to have a smooth convergence for some time. But eventually as we train for more number of epochs all 3 models attain almost same decreased loss.

1.4 Accuracies for test data

For beam_size = 10 we got

Execution Accuracy: 43.125!!

Exact Match Accuracy: 46.25!!

2 Bert2Seq Model tuned

2.1 Parameters

The best model obtained is for the following parameters:

- encoder hidden units: 768
- decoder hidden units: 768
- num_layers encoder: 1
- num_layers decoder: 1
- teacher forcing: 0.6
- learning_rate: 0.001
- Schedule step: 30
- num_epochs: 100

2.2 Training errors and Loss curves

2.2.1 Loss

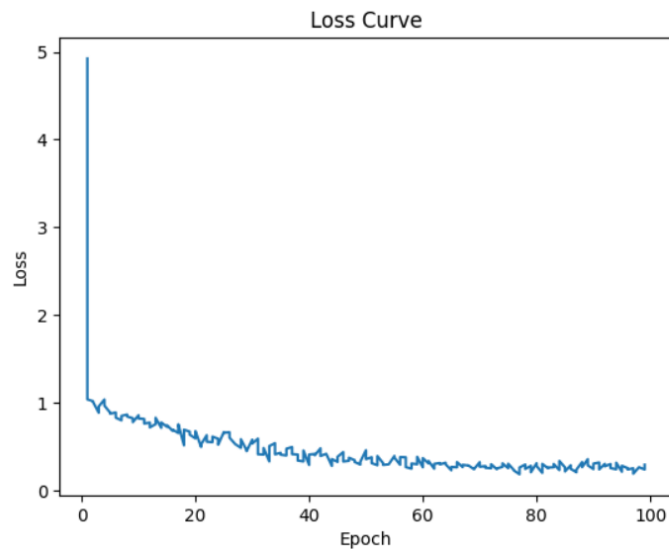


Figure 2: Plot of the effect of loss on the number of epochs of the training

The loss drops at a high rate with the start of training as there are many characters like brackets, commas, which can be learned by the model. Then the loss decreases slowly as the sequences are learnt and the learning rate is decreased gradually to ensure convergence.

2.3 Accuracies for test data

For beam_size = 1 we got
Execution Accuracy: 59.375!!
Exact Match Accuracy: 65.625!!
For beam_size = 10 we got
Execution Accuracy: 41.641!!
Exact Match Accuracy: 43.161!!
For beam_size = 20 we got
Execution Accuracy: 38.027!!
Exact Match Accuracy: 39.209!!

3 Seq2Seq Model with Attention

3.1 Parameters

The best model obtained is for the following parameters:

- **encoder hidden units:** 512
- **decoder hidden units:** 512
- **num_layers encoder:** 1
- **num_layers decoder:** 1
- **teacher forcing:** 0.3,0.6,0.9
- **learning_rate:** 0.001
- **Schedule step:** 30
- **num_epochs:** 100

3.2 Training errors and Loss curves

3.2.1 Loss

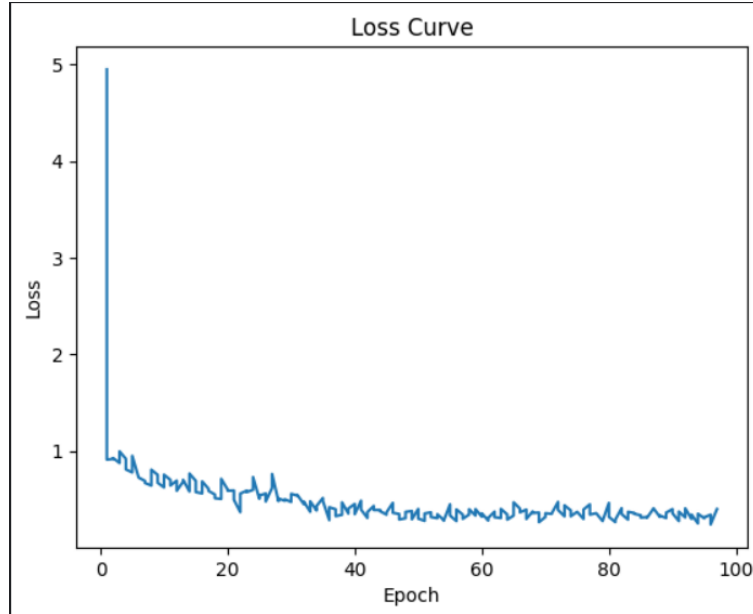


Figure 3: Plot of the effect of loss on the number of epochs of the training

The loss drops at a high rate with the start of training as there are many characters like brackets, commas, which can be learned by the model. Then the loss decreases slowly as the sequences are learnt and the learning rate is decreased gradually to ensure convergence.

3.3 Accuracies for test data

For beam_size = 10 we got

Execution Accuracy: 38.146!!

Exact Match Accuracy: 42.639!!

4 Seq2Seq Model

4.1 Parameters

The best model obtained is for the following parameters:

- **encoder hidden units:** 512
- **decoder hidden units:** 512
- **num_layers encoder:** 1
- **num_layers decoder:** 1
- **teacher forcing:** 0.6

- `learning_rate`: 0.001
- `Schedule step`: 30
- `num_epochs`: 100

4.2 Training errors and Loss curves

4.2.1 Loss

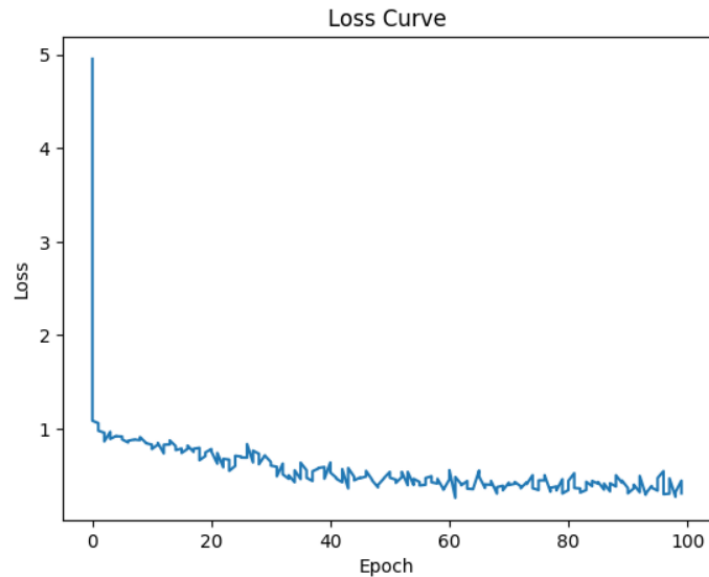


Figure 4: Plot of the effect of loss on the number of epochs of the training

The loss drops at a high rate with the start of training as there are many characters like brackets, commas, which can be learned by the model. Then the loss decreases slowly as the sequences are learnt and the learning rate is decreased gradually to ensure convergence.

4.3 Accuracies for test data

For `beam_size = 10` we got

Execution Accuracy: 28.72!!

Exact Match Accuracy: 32.987!!