Assignment 4 Writeup

Name: Rafayel Veziryan

Seq2Seq Results

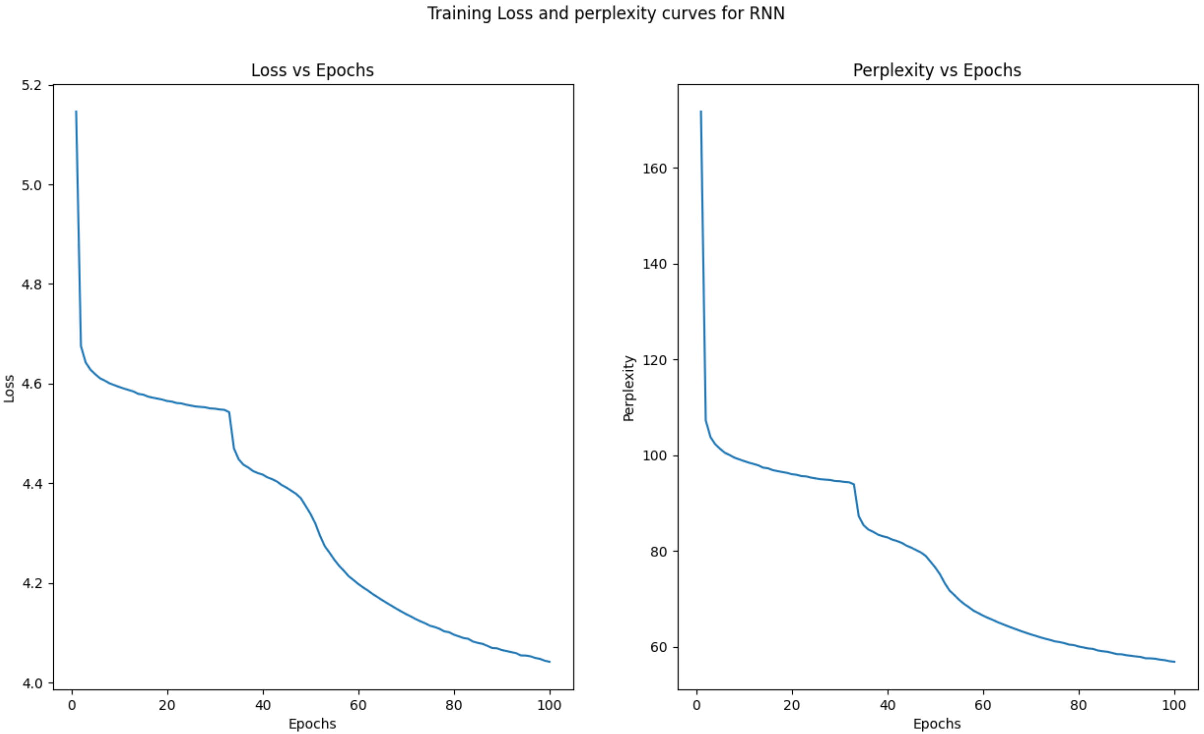
Table 1

Put your results from training before and after hyperparameter tuning here.

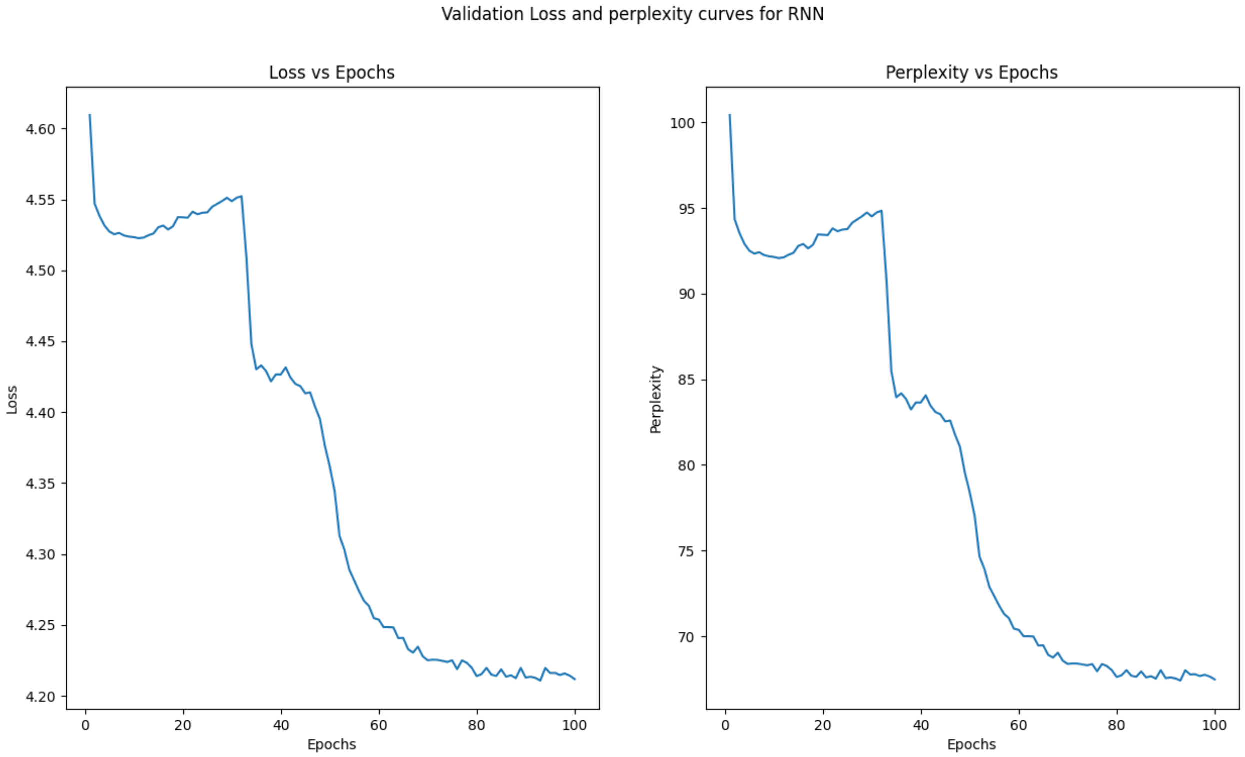
Results for default configuration using RNN		Results for default Configuration Using LSTM	
Training Loss	4.6057	Training Loss	4.3546
Training Perplexity	100.052	Training Perplexity	77.8356
Validation Loss	4.5238	Validation Loss	4.2345
Validation Perplexity	92.186	Validation Perplexity	69.0297
Result for your Best Model using RNN after hyperparameter tuning		**Resut for your Best Model using LSTM after hyperparameter tuning**	
Training Loss	4.04	Training Loss	4.01
Training Perplexity	56.9	Training Perplexity	55.45
Validation Loss	4.21	Validation Loss	4.2
Validation Perplexity	67.48	Validation Perplexity	67
Your best model configuration for RNN after hyperparameter tuning		**Your best model configuration for LSTM after hyperparameter tuning**	
Batch size = 64, LR = 0.001, OR Epochs = 100		Batch size = 64, LR = 0.001, Epochs = 100	

Seq2Seq Curves

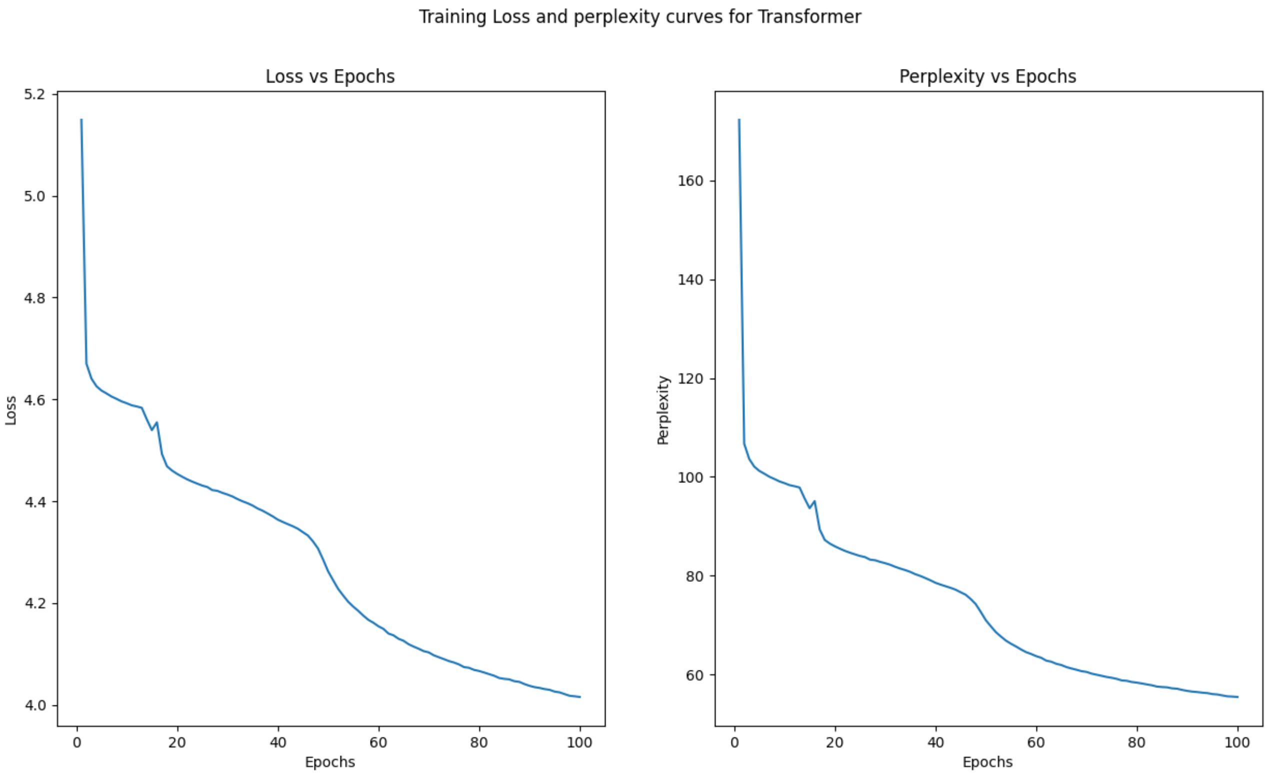
Put the plots for loss and perplexity curves (training & validation) for your configuration with default setting and for your best model here. Use additional slides as necessary.



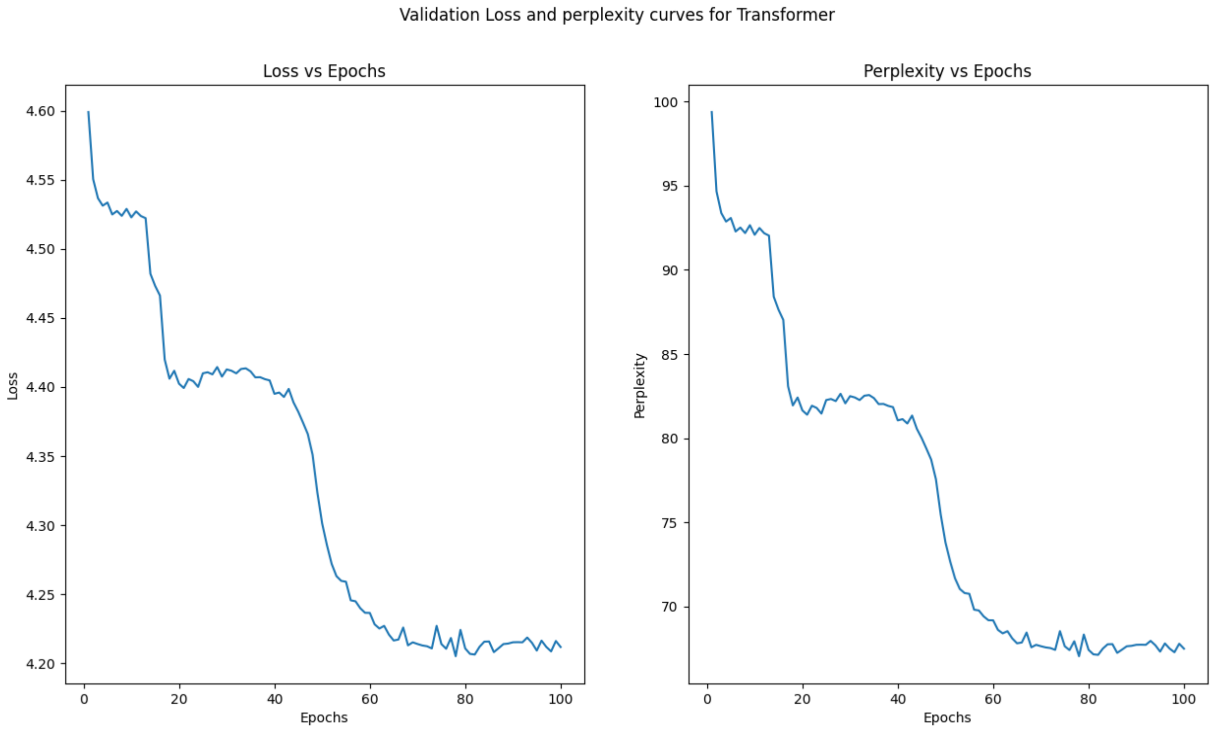
RNN Validation loss and perplexity curves



LSTM training loss and perplexity curves



LSTM validation loss and perplexity curves



Seq2Seq Explanation

Explain what you did here and why you did it to improve your model performance. Compare and explain the differences when using LSTM vs RNN. You may use another slide if needed.

I implemented seq2seq model with RNN and LSTM cells. During model improvement used grid search method for finding better hyperparameters such as learning rate, batch size and number of epochs, to affect model efficiency, we also can experiment taking new optimizer and learning rate scheduler, we should research which optimizer and learning rate scheduler can match with our task. The first obvious difference between RNN and LSTM is RNN has simple recurrent connection to itself, while LSTM has more complex architecture with input, forgot and ouptut gates. RNNs have fewer parameters as compared with LSTM. Using LSTM we can take much more information thanks to their architecture and more parameters.

Transformer Results

Table 2

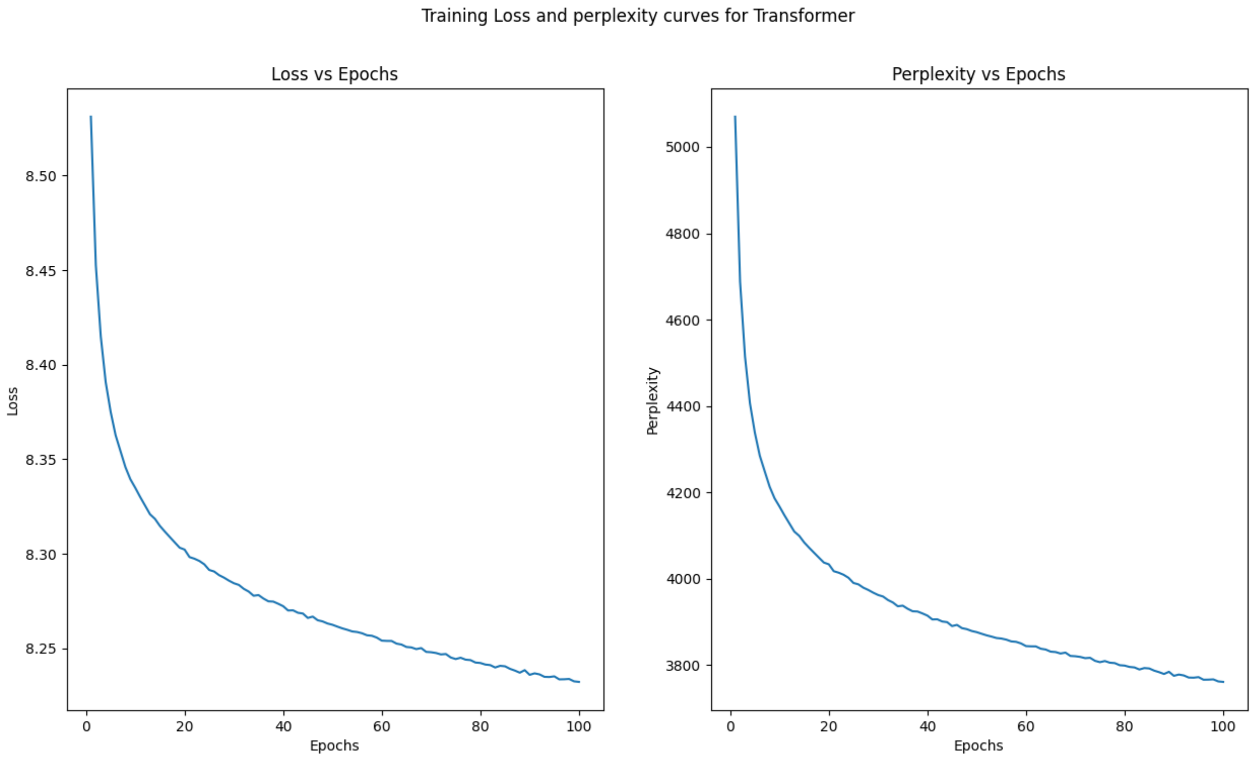
Put your results from training before and after hyperparameter tuning here.

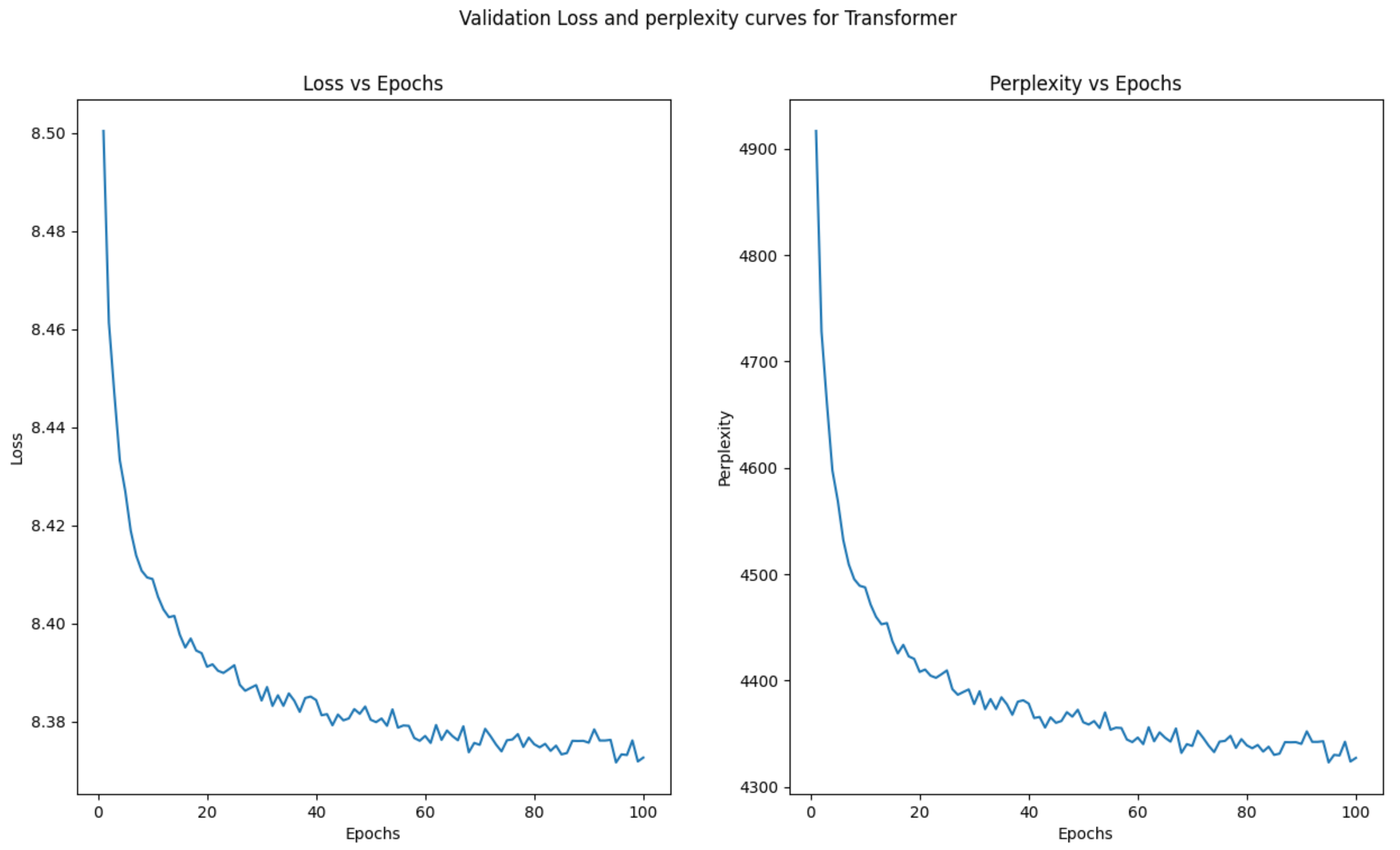
Results for default configuration with sine/cosine encoding			
Training Loss	8.3351	Validation Loss	8.4074
Training Perplexity	4167.62	Validation Perplexity	4473.809
Result for your Best Model			
Training Loss	8.232	Validation Loss	8.37
Training Perplexity	3760.43	Validation Perplexity	4323
Your best model configuration after hyperparameter tuning			
Batch size = 32, learning rate = 0.001, epoch number = 100			

Results for default configuration with learnable positional encoding			
Training Loss	8.3347	Validation Loss	8.411
Training Perplexity	4165.9118	Validation Perplexity	4496.3348

Transformer Curves

Put the plots for loss and perplexity curves (training & validation) for your configuration with default setting and for your best model here. You may use additional slides if needed.





Validation Loss and perplexity curves for Transformer
Loss vs Epochs
Loss
Epochs
8.50
8.48
8.46
8.44
8.42
8.40
8.38
0
20
40
60
80
100
Perplexity vs Epochs
Perplexity
Epochs
4900
4800
4700
4600
4500
4400
4300
0
20
40
60
80
100

Transformer Explanation

Explain what you did here and why you did it to improve your model performance. You may use another slide if needed.

I implemented transformer encoder block with sine/cosine and learnable positional encoding for machine translation task. For this model I also used grid search method to find much better hyperparameters such as batch size, learning rate and number of epochs. To improve model efficiency we also can modify optimizer and learning rate scheduler.

Transformer Translation Results

Table 3

Put translation results for your best model (1st 9 sentences) here

Input sentence	Back translation
'<sos>' 'a' 'man' 'cooking' 'burgers' 'on' 'a' 'black' 'grill' '.' <eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'man' 'guy' 'cake' 'sheet' 'sheet' 'cooking' 'leaning' 'leans' 'red' 'big' 'grill' 'black' 'grill' 'tree' 'black' 'black' '.' 'black' '<eos>' 'a'
'<sos>' 'a' 'man' 'and' 'woman' 'fishing' 'at' 'the' 'beach' '.''<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>''<pad>'	'man' 'and' 'and' 'glasses' 'wife' 'woman' 'fishing' 'fishing' 'at' 'at' 'beach' 'beach' 'beach' 'beach' 'beach' 'beach' 'beach' 'beach' 'beach' 'beach'
'<sos>' 'a' 'man' 'in' 'a' 'harness' 'climbing' 'a' 'rock' 'wall' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>''<pad>'	'man' 'machine' 'hole' 'machine' 'rock' 'climbing' 'climbing' 'climbing' 'climbing' 'wall' 'face' 'stone' 'wall' 'wall' 'rock' 'rock' 'wall' 'wall' 'stone' 'rock'
'<sos>' 'a' 'cute' 'baby' 'is' 'smiling' 'at' 'another' 'child' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' <pad>'	'another' 'another' 'friend' 'baby' 'smiling' 'smiles' 'another' 'child' 'who' 'child' 'child' 'child' 'child' 'child' 'child' 'another' 'another' 'another' 'another' 'another'
'<sos>' 'a' 'female' 'playing' 'a' 'song' 'on' 'her' 'violin' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'female' 'happy' 'sniffing' 'plays' 'plays' 'song' 'sand' 'her' 'violin' 'violin' 'violin' 'violin' 'violin' 'her' 'her' 'her' 'her' 'her' 'her' 'her'
'<sos>' 'a' 'person' 'on' 'a' 'snowmobile' 'in' 'mid' 'jump' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'person' 'person' 'bed' 'onto' 'bicycle' 'bicycle' 'mid' 'middle' 'jump' 'middle' 'jump' 'jump' 'the' 'the' 'the' 'them' 'grass' 'jump' 'them' 'middle'
'<sos>' 'three' 'men' 'competing' 'in' 'a' 'hurdle' 'race' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'men' '3' '3' 'perform' 'competing' 'competing' 'race' 'well' 'competing' 'competing' 'match' 'match' 'match' 'other' 'other' 'match' 'match' 'at' 'match' 'other'
'<sos>' 'people' 'play' 'in' 'a' 'fountain' 'at' 'twilight' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'people' 'fountain' 'playing' 'shorts' 'surfing' 'surfing' 'shaving' 'in' 'fountain' 'fountain' 'fountain' 'fountain' 'a' 'fountain' 'fountain' 'in' '.' '.' '<eos>' 'the'

LSTM Translation Results

Table 4

Put translation results for your best model (1st 9 sentences) here

Input sentence	Back translation
“<sos>' 'a' 'man' 'cooking' 'burgers' 'on' 'a' 'black' 'grill' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'a' 'man' 'is' 'a' 'a' '<unk>' 'a' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'a' 'man' 'and' 'woman' 'fishing' 'at' 'the' 'beach' '.''<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>''<pad>'	'<sos>' 'a' 'man' 'and' 'a' 'are' 'are' 'on' 'the' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'a' 'man' 'in' 'a' 'harness' 'climbing' 'a' 'rock' 'wall' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>''<pad>'	'<sos>' 'man' 'man' 'in' 'climbing' 'climbing' 'climbing' 'a' 'rock' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'a' 'cute' 'baby' 'is' 'smiling' 'at' 'another' 'child' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' <pad>'	'<sos>' 'a' 'smiling' 'baby' 'is' 'a' 'a' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'a' 'female' 'playing' 'a' 'song' 'on' 'her' 'violin' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'a' 'woman' 'playing' 'a' 'guitar' 'guitar' 'a' 'the' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'a' 'person' 'on' 'a' 'snowmobile' 'in' 'mid' 'jump' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'a' 'person' 'on' 'a' 'a' 'a' 'a' 'a' 'a' 'a' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'three' 'men' 'competing' 'in' 'a' 'hurdle' 'race' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'three' 'men' 'are' 'a' 'on' 'a' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'people' 'play' 'in' 'a' 'fountain' 'at' 'twilight' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'people' 'are' 'in' 'in' 'a' 'in' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'
'<sos>' 'the' 'three' 'children' 'are' 'in' 'a' 'cage' '.' '<eos>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>' '<pad>'	'<sos>' 'four' 'young' 'are' 'are' 'in' 'a' '.' '.' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>' '<eos>'

Compare LSTM to Transformer

Compare your LSTM results to your Transformer Results both quantitatively and qualitatively and explain the differences.

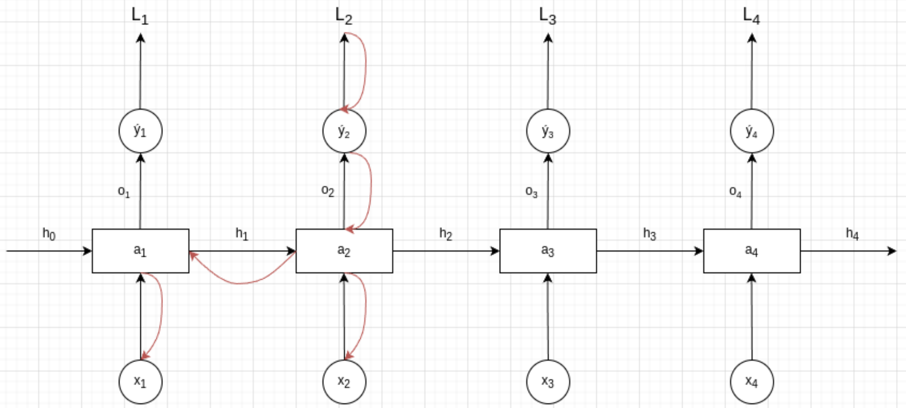
Surprisingly, from LSTM and Transformer results appears that LSTM works much better than Transformer. This phenomenon also considered for such examples, which are introduced in last slides. Validation loss for Transformer is 8.37 with two times more as compared with LSTM. This mean, we have to find problems, because Transformer has much more deeper architecture and parameters than LSTM.

Best results model parameters saved in [Drive]

Theory question

Add additional slides as necessary for your answer

This file is attached with other files, called diagram_L2_backprop_drawio_Rafayel_Veziryan.pdf



$a_t = Ux_t + Wh_{t-1} + b$

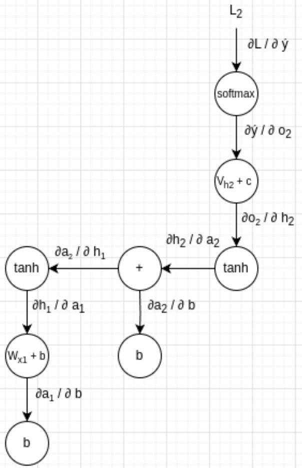
$h_t = \tanh(a_t)$

$o_t = Vh_t + c$

$\acute{y}t = \mathrm{Softmax}(o_t)$

$L_t = CE(\acute{y}t , yt)$

$f(x) = \tanh(x) \longrightarrow f'(x) = 1 - \tanh^2(x)$



$\partial L / \partial b = (\partial L / \partial \acute{y}) * (\partial \acute{y} / \partial o_2) * (\partial o_2 / \partial h_2) * (\partial h_2 / \partial a_2)*((\partial a_2 / \partial b) + (\partial a_2 / \partial h_1) * (\partial h_1 / \partial a_1) * (\partial a_1 / \partial b)) =$

$= (\acute{y}_2 - y_2)*V*(1-\tanh^2 (a_2)(1 + W(1 - \tanh^2 (a_1))$