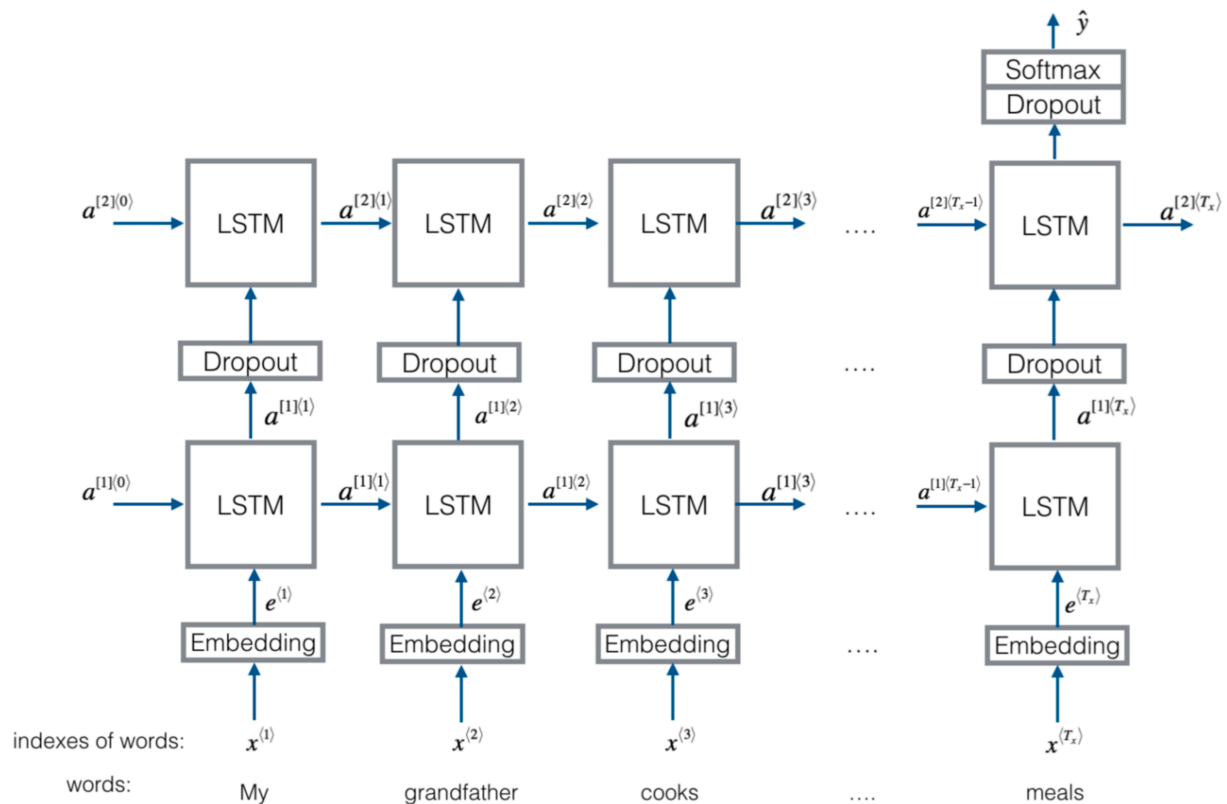


HW5 — Sentence Level Sentiment Classification

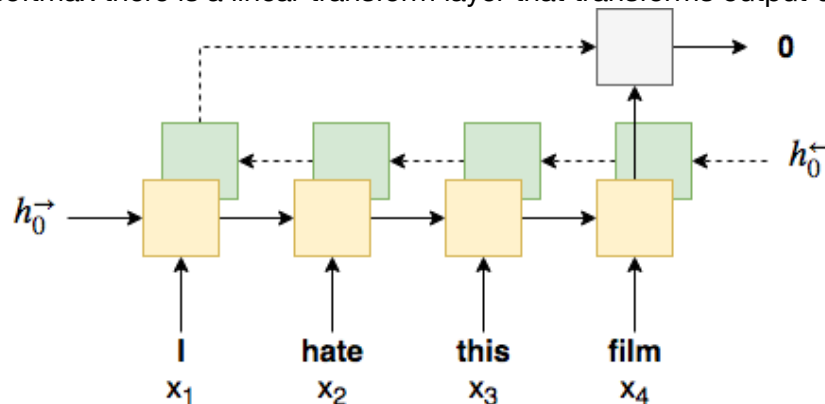
1. Model Architecture

Sentence-level sentiment classification was performed on a 2-layer bidirectional lstm with dropout between layer outputs.

The architecture used is shown below:



*before the softmax there is a linear transform layer that transforms output dimension to 5



*depiction of the bi-directionally of one layer

outputs from both directions are *concatenated* to form the output of the lstm

source: <https://github.com/bentrevett/pytorch-sentiment-analysis>

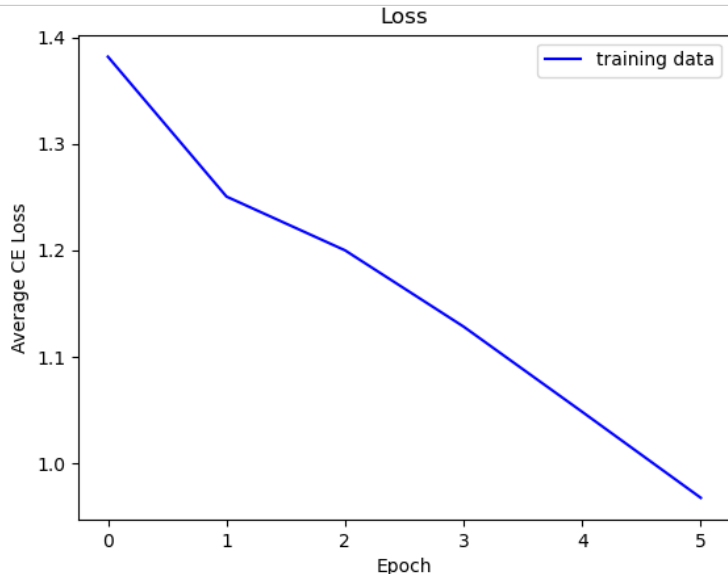
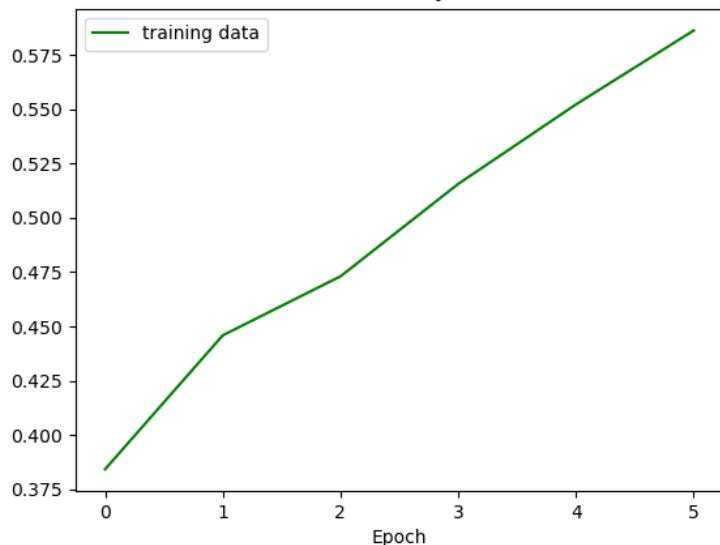
The model parameters are listed below:

- embedding size = 300 (taken from provided vector.txt file)
- hidden dimension = 256
- number of lstm layers = 2
- bidirectionality = True
- dropout = 0.5

The training parameters used are listed below:

- optimizer = Adam
- batch_size = 64
- max_epoch = 6
- learning_rate = 0.001

1. Performance

	Training Data	Test Data
Loss	0.988 	1.3365
Accuracy	0.5719 	0.4417

2. Conclusions

The model is extremely prone to overfitting the training data. I have also trained the same model for 10 epochs, where it can reach training accuracies of up to 90%. However, the test accuracy performance decreases to only 40%. Increasing the dropout/regularization can greatly prevent the model from overfitting but will not completely eliminate this issue.

I also encountered another issue when training the model using the classical stochastic gradient descent with momentum. The model seems to converge with a training accuracy of only ~27% and does not learn any further. However, switching to Adam optimizer solved this issue. This seems to imply that sgd with momentum caused a convergence at a local (rather than a global) minima. Lesson learned... I will solely be using Adam from now on.