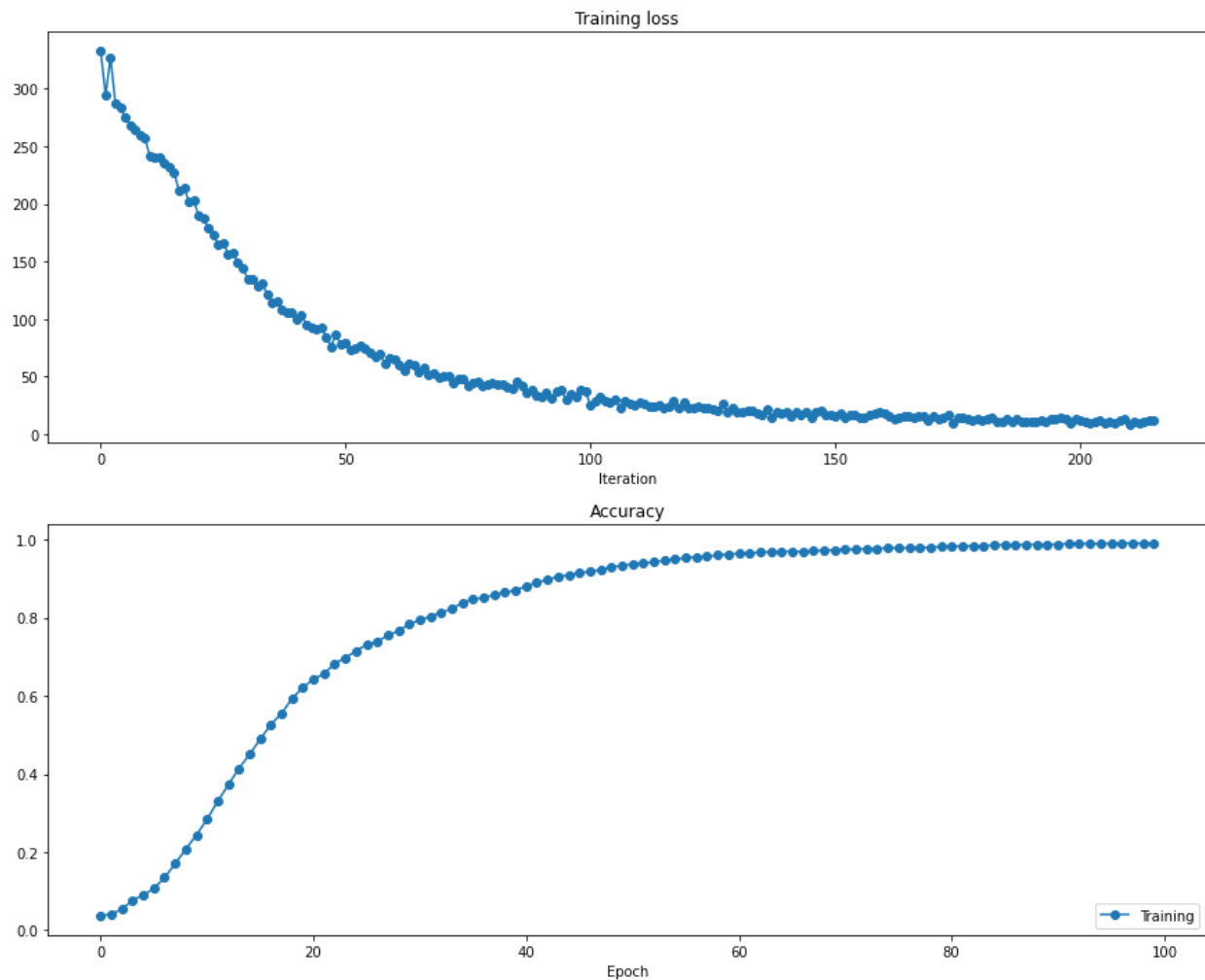


CSCI566-Deep Learning and Its Applications

Problem 2 Solutions

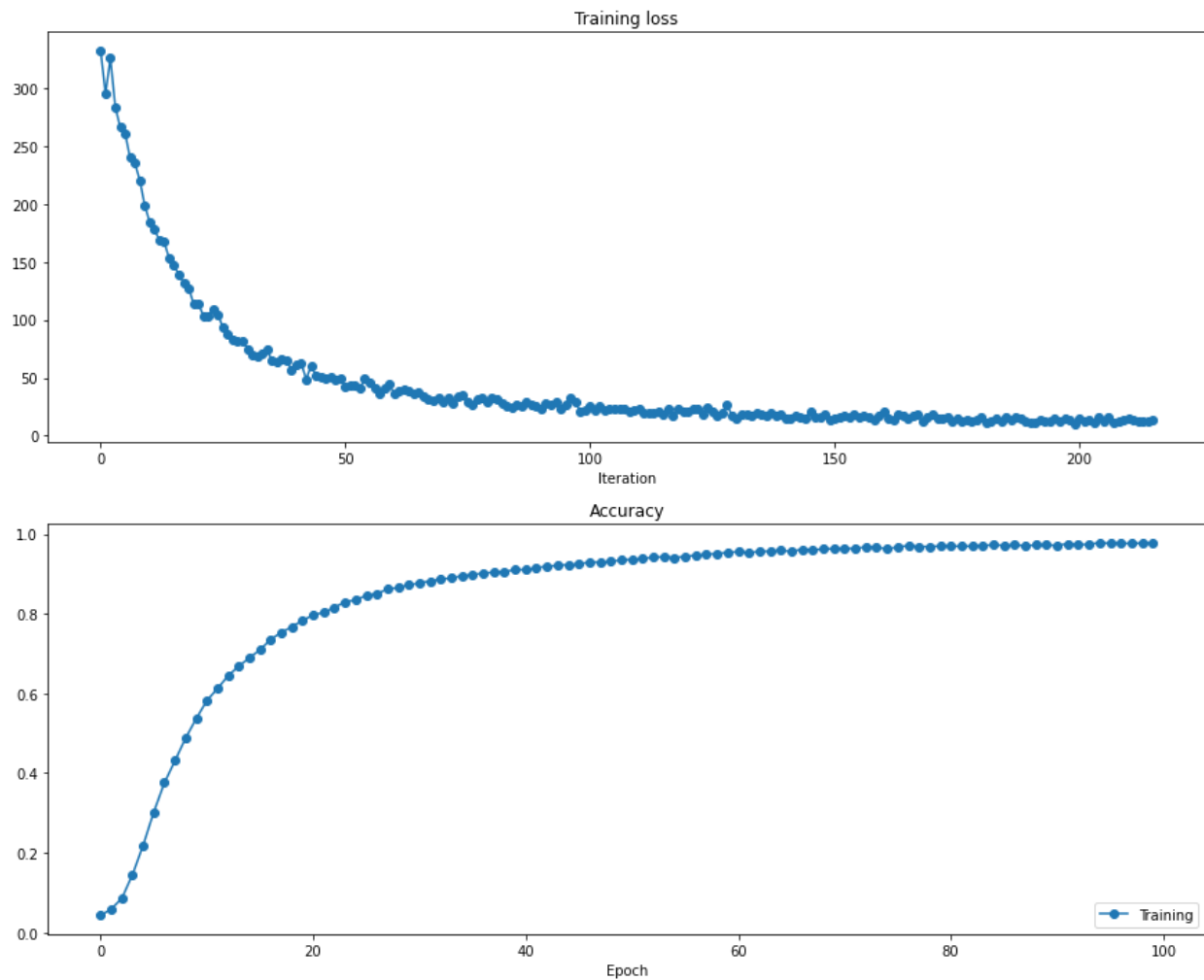
1. Training loss / accuracy curves for vanilla RNN and LSTM training

LSTM Training loss / accuracy curves



best performance 99.12402028584602%

Vanilla RNN Training loss / accuracy curves



best performance 97.83310281235592%

The best accuracy attained by Vanilla RNN is 97.60 percent and LSTM is 99.124 percent. LSTM is performing better. But for a test length of 40, there is not much of a difference in the output text generated.

2. Sample text generation from a trained model

Vanilla RNN, timestep=50, test length=40

she had begun to think that she was now only ten inches high, and her face brightened up at the thought that she was now the right size for going through the little door into that lovely garden. first, however,

LSTM, timestep=50, test length=40

she went back to the table, half hoping she might find another key on it, or at any rate a book of rules for shutting people up like telescopes: this time she found a little bottle on it, [‘which certainly

But for a test length of 40, there is not much of a difference in the output text generated by both the models

Vanilla RNN, timestep=100 and test length=200

she got to the door, she was coming to, but it was too dark to see anything; then she looked at the sides of the well, and noticed that they were filled with cupboards and book-shelves; here and there she saw maps and pictures hung upon pegs. she took down a jar from one of the shelves as she passed; it was labelled ‘orange marmalade’, but to her great disappointment it was empty: she had plenty of time as she went down to look about her and to wonder what was going to happen next. first, she tried to look down and make out what she was coming to, but it was too dark to see anything; then she looked at the sides of the well, and noticed that they were filled with cupboards and book-shelves; here and there she saw maps and pictures hung upon pegs. she took down a jar from one of the shelves as she passed; it was labelled ‘orange marmalade’, but to her great disappointment it was empty: she had plenty of time as she went down to look about her and to wonder what was going to happen next. first, she tried to look

LSTM, time step=100, test length=200

she was now the thought that she was playing against herself, for this curious child was very fond of pretending to be two people. ‘but it’s no use now,’ thought poor alice, ‘to pretend to be two people! why, there’s hardly enough of me left to make one respectable person!’ soon her eye fell on a little glass box that was lying under the table: she opened it, and found in it a very small cake, on which the words ‘eat me’ were beautifully marked in currants. ‘well, i’ll eat it,’ said alice, ‘and if it makes me grow larger, i can reach the key; and if it makes me grow smaller, i can creep under the door; so either way i’ll get into the garden, and i don’t care which happens!’ she ate a little bit, and said anxiously to herself, ‘which way? which way?’, holding her hand on the top of her head to feel which way it was growing, and she was dozing off, and had just begun to dr

eam that she was walking hand in hand with dinah, and saying to her very earnestly, 'now, dinah, tell me the truth: did you ever eat

Observations: But as the test length and the time steps are increased, we can see that LSTM performs better in generating long text sentences without much repetition compared to Vanilla RNN.

3. Answers to inline questions about recurrent net behavior

Ans: We can observe that with the initial condition provided in the problem statement with learning rate set to $1e-03$ the training accuracy obtained by LSTM is more than that of Vanilla RNN. But as the test length and the time steps are increased, we can see that LSTM performs better in generating long text sentences without much repetition compared to Vanilla RNN. Vanishing gradient problem increasing with longer sequences from past timesteps, as the network is not able to learn. Thus, LSTM was designed to handle vanishing gradients and it is clearly seen that it performs better at longer text generation compared to Vanilla RNN.

Limitations in training recurrent language models is dealing with vanishing gradients and exploding gradients. In case of vanishing gradients, if the gradients obtained are very small, as these gradients traverse through the network the gradients get smaller and smaller due to chain rule and thus the gradient vanishes and the network does not learn. In the exploding gradients issue the gradients explode thus, the model can never reach a optimal value.