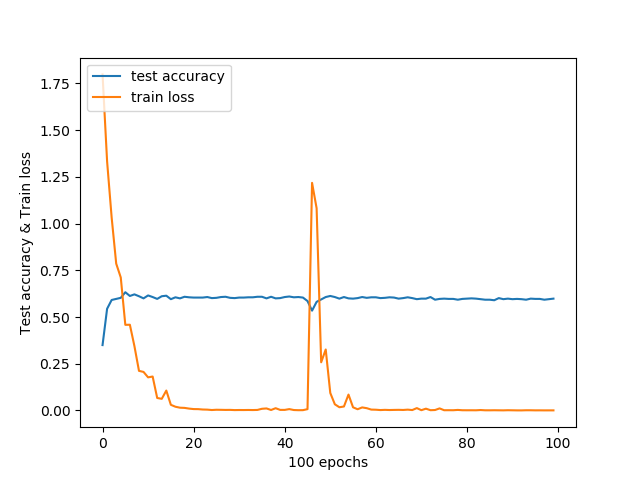
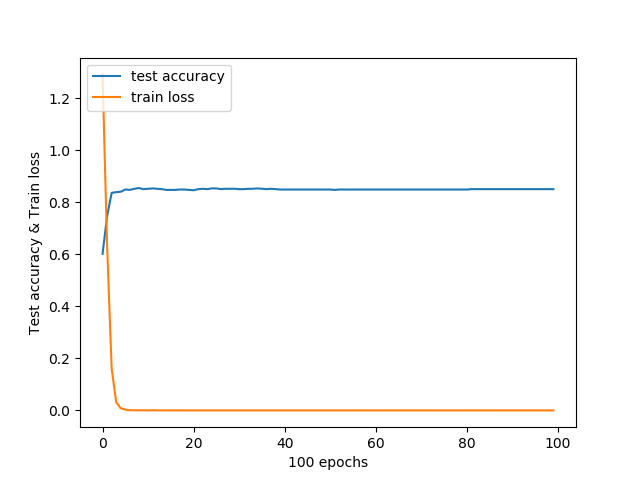
**Part B: Text classification**

**1. Plot the entropy cost on the training data and the accuracy on the testing data against training**

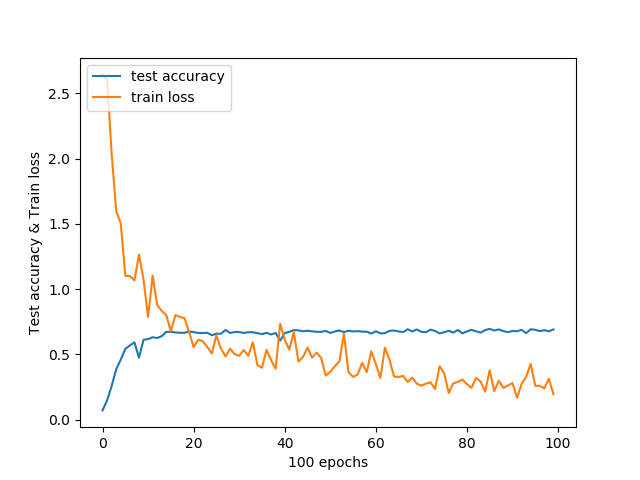
**epochs.**

****

* 1. **2. Plot the entropy cost on training data and the accuracy on testing data against training epochs.**
  2. ****

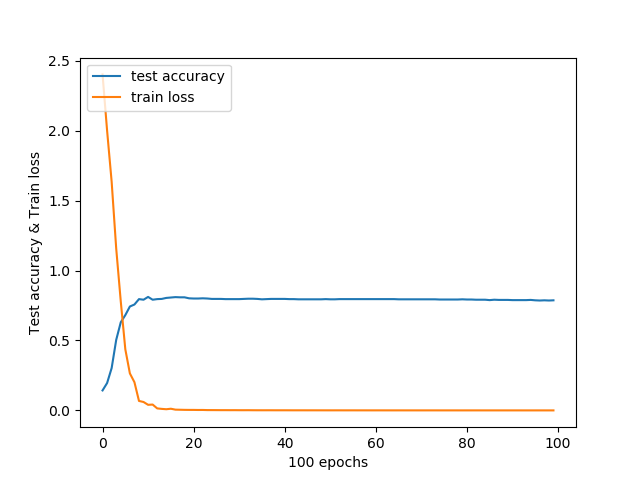
**3. Design a Character RNN Classifier that receives character ids and classify the input. The RNN is GRU layer and has a hidden-layer size of 20.**

**Plot the entropy cost on training data and the accuracy on testing data against training epochs.**

****

**4. Design a word RNN classifier that receives word ids and classify the input. The RNN is GRU layer and has a hidden-layer size of 20. Pass the inputs through an embedding layer of size 20 before feeding to the RNN.**

**Plot the entropy on training data and the accuracy on testing data versus training epochs.**

****

**5. Compare the test accuracies and the running times of the networks implemented in parts (1) – (4).**

**Experiment with adding dropout to the layers of networks in parts (1) – (4), and report the test accuracies. Compare and comment on the accuracies of the networks with/without dropout.**

|  |  |  |
| --- | --- | --- |
| **Model** | **Test Accuracy** | **Running Times (seconds)** |
| Q1 without dropouts | 0.5985714 | 1106 |
| Q2 without dropouts | 0.85 | 183 |
| Q3 without dropouts | 0.69142854 | 423 |
| Q4 without dropouts | 0.7871429 | 141 |
| Q1 with dropouts | 0.6414286 | 1030 |
| Q2 with dropouts | 0.9028571 | 186 |
| Q3 with dropouts | 0.66 | 425 |
| Q4 with dropouts | 0.89 | 138 |

Run times:

1. Comparing models with & without dropouts, both runtimes are similar.
2. Character processing models takes substantially longer time to train than word processing models.
3. RNN models are substantially faster to train compared to CNN models.

Accuracies:

1. For the most part, models with dropouts have better accuracies than models without dropouts.
2. Word processing models are more accurate than character processing models
3. CNN & RNN models have similar performance for accuracies.

In this experiment, both CNN & RNN models have similar performance. However, RNN models have a slight advantage over CNN in running times.

**6. For RNN networks implemented in (3) and (4), perform the following experiments with the aim of improving performances, compare the accuracies and report your findings:**

**a. Replace the GRU layer with (i) a vanilla RNN layer and (ii) a LSTM layer**

**b. Increase the number of RNN layers to 2 layers**

**c. Add gradient clipping to RNN training with clipping threshold = 2.**

|  |  |
| --- | --- |
| **Model** | **Test Accuracy** |
| Q3 | 0.69142854 |
| Q4 | 0.7871429 |
| Q3 with Vanilla RNN layer |  |
| Q4 with Vanilla RNN layer |  |
| Q3 with LSTM layer |  |
| Q4 with LSTM layer |  |
| Q3 with 2 RNN layers |  |
| Q4 with 2 RNN layers |  |
| Q3 with gradient clipping | 0.62857145 |
| Q4 with gradient clipping | 0.87142855 |