**Part B: Text classification**

**1. Design a Character CNN Classifier that receives character ids and classifies the input. The CNN has two convolution and pooling layers:**

**• A convolution layer 𝐶1 of 10 filters of window size 20x256, VALID padding, and ReLU neurons. A max pooling layer 𝑆1 with a pooling window of size 4x4, with stride = 2, and padding = 'SAME'.**

**• A convolution layer 𝐶2 of 10 filters of window size 20x1, VALID padding, and ReLU neurons. A max pooling layer 𝑆2with a pooling window of size 4x4, with stride = 2 and padding = 'SAME'.**

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

**epochs.**

* 1. **2. Design a Word CNN Classifier that receives word ids and classifies the input. Pass the inputs through an embedding layer of size 20 before feeding to the CNN. The CNN has two convolution and pooling layers with the following characteristics:**
  2. **• A convolution layer 𝐶1 of 10 filters of window size 20x20, VALID padding, and ReLU neurons. A max pooling layer 𝑆1 with a pooling window of size 4x4, with stride = 2 and padding = 'SAME'.**
  3. **• A convolution layer 𝐶2 of 10 filters of window size 20x1, , VALID padding, and ReLU neurons. A max pooling layer 𝑆2with a pooling window of size 4x4, with stride = 2 and padding = 'SAME'.**

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

**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.**

**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.**