CS512: Advanced Machine Learning.

Assignment 3: Adversarial Training on Sequence Classification

Garima Gupta: ggupta22@uic.edu

Sai Teja Karnati: skarna3@uic.edu

Shubham Singh: ssing57@uic.edu

Wangfei Wang: wwang75@uic.edu

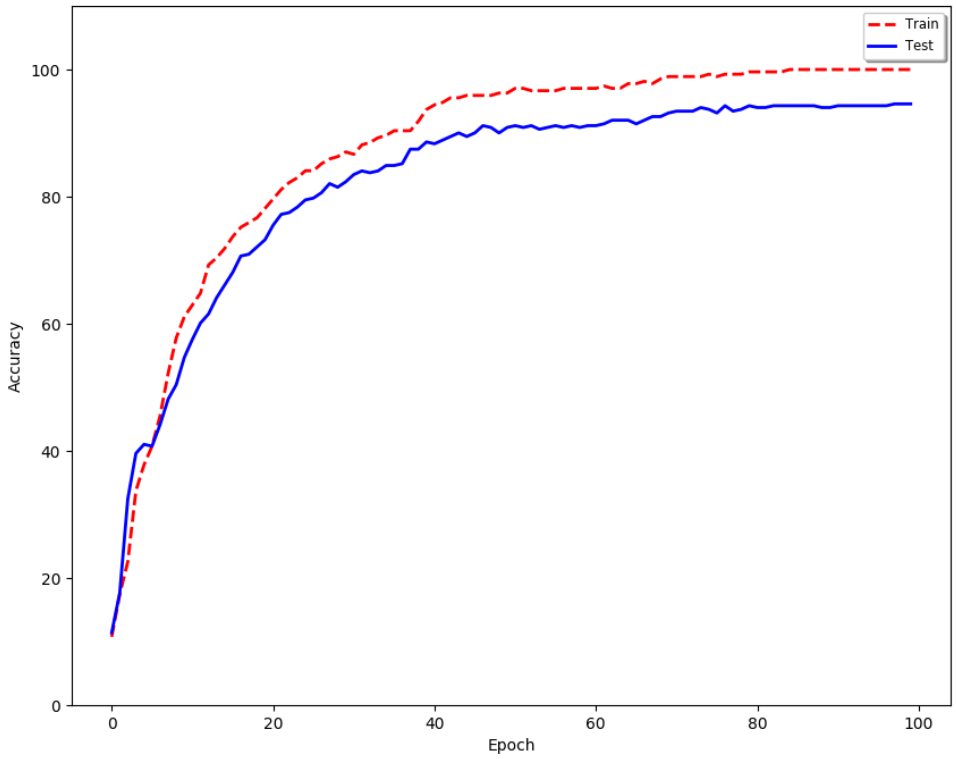
April 13, 2020

1 Introduction

2 (15 points) Training the Basic Model

Hyperparameters values:

`batch_size = 27`, `hidden_size = 10`, `basic_epoch = 100`, `out_channels = 64`, `kernel_size = 10`, `stride = 3`, `lr = 1e-3` (learning rate), `weight_decay = 1e-3`.



3 (10 points) Save and Load Pretrained Model

See code in `training.py`. In the code, we commented this part as `Part 3, Save and Load model`.

4 (25 points) Adversarial Training as Regularization

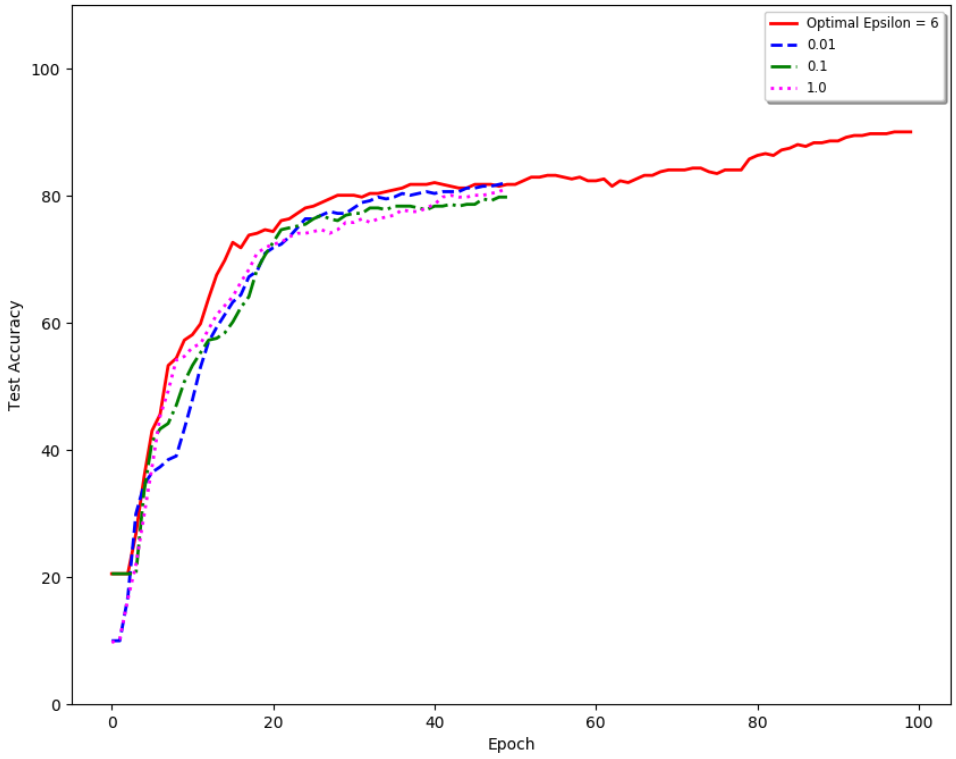
a **(10 points)** See the `compute_perturbation` function in `training.py`.

b **(5 points)** See the branch `mode` = 'AdvLSTM' in `LSTMClassifier` in `Classifier.py`.

c **(10 points)**

Among the ϵ's we have tried ($\epsilon = [0, 2, 4, 6, 8, 10, 0.001, 0.01, 0.1, 1, 10, 100, 1000]$), $\epsilon = 6$ gives the optimal performance at the end of 100 epochs. The other hyperparameters were set the same as those in the basic model.

As shown in the figure above, the performance of the model changes slightly with the change of ϵ, meaning our model is pretty robust to disturbance. At the end of epoch 50, $\epsilon = 0.01$ seems to give the best test accuracy among $\epsilon = [0.01, 0.1, 1]$. But again, the test accuracies are pretty similar in the set of ϵ's we have tried.



5 (40 points) Adversarial Training as Proximal Mapping

6 (10 points) Dropout and Batch Normalization