CSA 250 : Deep Learning Project III
Report

Nilesh Kande

May 17, 2020

1 Model link:

Unable to upload model on github as size exceeds 100 MB.
Link: https://drive.google.com/open?id=1l7IsCH8OH3vxljnw5RVdzHVvlCrOpBCm

2 Accuracy:

TF-IDF Network TRAIN Accuracy: 51.4
RNN Network TRAIN Accuracy: 74.74 RNN Network Validation Accuracy: 65.48

TF-IDF Network TEST Accuracy: 50.14
RNN Network TEST Accuracy: 94.10

3 Dataset

Dataset was accessed using downloaded data from stanford site. It was pre-processed using in-built libraries. Dataset is available in three sections as train set, dev(validation) set and test set.
Train example : 550152
Validation example : 10000
Test example : 10000

4 TF-IDF Network :

4.1 Model Architecture:

- First, a document was created using all words from entire dataset corpus.
- Conversion to lower case, stop words removal, stemming and tokenisation was done using nltk libraries (stopword and tokenization)
- TF-IDF formulation was applied on this corpus
- Then, in-built library of sklearn.LogisticRegressionCV was used for training to check how linear regression will perform on data having temporal dependency.

5 RNN Network:

5.1 Model Architecture:

- First, data was passed through an embedding layer.
- Then, output of that layer was passed to the RNN layer.
- It was then Passed through the Multi-layer perceptron layers, of which first layer had ReLu as it's activation function and second layer had outputs as equal to the number of classes which is 3.

5.1.1 Parameters:

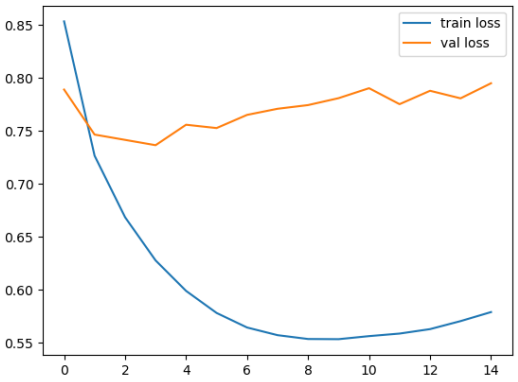
optimizer : Adam
learning rate : 0.0005
epochs = 15
batch size = 32
Monitoring parameter to save the model : Validation loss

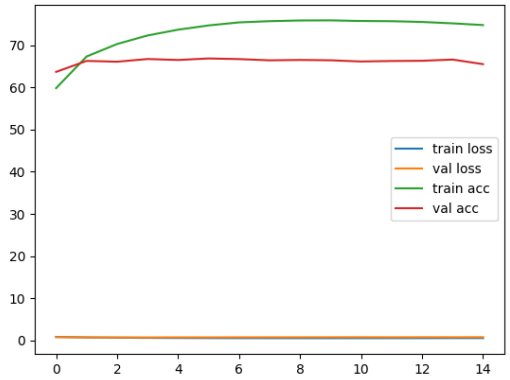
5.2 Model Selection:

Model Architecture was fixed from the beginning as stated in precious sub-section. Model was mainly fine-tuned over what should be the embedding size for the word vector.
Various embedding sizes were tried on and the one with the best result are submitted. Results for various embedding sizes are reported in accuracy section.
Best model was saved using validation loss. It was checked over validation loss as well as validation accuracy.

5.3 training/validation Loss Graph:

This graphs shows validation and training loss against number of epochs and validation and training accuracy against number of epochs respectively.





6 Accuracy:

Embeeding size - 100, RNN Network Accuracy: 60.57 Embeeding size - 800, RNN Network Accuracy: 62.50 Embeeding size - 500, RNN Network Accuracy: 66.71

7 Experiments:

Framework used was pytorch
Training is done on 12 GB GPU.
Testing requires CPU only.