

Neural Networks and Deep Learning: Lab Assignment 3

Summary: In this assignment, you will demonstrate you understand how to train, evaluate, and analyze recurrent neural networks (RNNs). Your submission should include one PDF file with two separate, self-contained parts: one being a report describing your methods, results, and analysis and the other part showing your code.

1. Impact of RNN Architecture [15 points]:

(a) Design and conduct your experiment (Code)

- Load a dataset for spam classification. If you are seeking assistance for how to find a dataset, one good option is to revisit the TA's programming tutorials. You are welcome to pre-process the dataset.
- If the dataset you choose does not already come in a train/test split, then divide it into a 70/30 train/test split of the dataset.
- Train three models of your design choice which are identical except that they are based on a vanilla RNN, LSTM, and GRU respectively. Train all three models identically; e.g., use the same input format, hyperparameters, and training approach.
- For each final model, compute the precision and recall on the test set. Then, perform fine-grained analysis with respect to input length by (1) dividing your test set into short, medium, and long inputs that are roughly three equal-sized sets based on the number of words observed in the test set examples and then (2) for each model, report the precision and recall values independently for short, medium, and long inputs.

(b) Report your methods, results, and analysis (Report)

- Describe the methods you used for your experiment such that the reader could reproduce your experiments. This should include a discussion of the dataset (e.g., source? number of examples?), what neural network architectures and hyperparameters were used to train all the models, and what type of hardware was used during training.
- For all models, report the precision and recall overall as well as with respect to the fine-grained analysis based on input length.
- Discuss your analysis of general trends that emerge from your results. Your discussion should consist of 2-4 paragraphs. To format each paragraph, please first identify one general trend observed from the results and then offer insights/speculations into **why** you think the trend/results may occur (regardless of whether you deem the results good or bad). Possible trends to consider include: Did a certain type of RNN architecture lead to better results? Are there performance trends across the three models with respect to how they handle short vs medium vs long inputs? What, if any, insights are gained by looking at the different evaluation approaches (i.e., precision, recall)?

2. Impact of Pretrained Word Embedding [10 points]:

(a) Design and conduct your experiment (Code)

- Use the same dataset with splits from the previous problem.
- Train two RNN models of your design (e.g., choose vanilla RNN vs LSTM vs GRU, number of units per hidden layer, number of layers, activation functions) where the only difference is type of pretrained word embedding used to represent the input (e.g., word2vec, GloVe). Train both models identically; e.g., use the same input format, hyperparameters, and training approach.
- For each trained model, produce a confusion matrix and compute the precision and recall.

(b) Report your methods, results, and analysis (Report)

- Describe the methods you used for your experiment such that the reader could reproduce your experiments. This should include a discussion of the dataset (e.g., source? number of examples?), what neural network architectures and hyperparameters were used to train all the models, and what type of hardware was used during training.
- For both models, report the confusion matrices, precisions, and recalls.
- Discuss what general trends emerge from your results. Some suggestions for topics to discuss are listed here. Did a certain type of word embedding lead to better results? What, if any, insights are gained by looking at the different evaluation approaches (i.e., confusion matrix, precision, recall)? How would you expect these word embeddings to compare to a baseline model that uses the tokenized representation as input instead?

How to Submit Lab Assignment 3: Please submit a pdf named with your first and last name; i.e., `firstname_lastname.pdf`. A successful submission will consist of two self-contained, separate contributions. First, it should include a report describing all methods, results, and analysis (i.e., portions indicated by “Report”) **as the first part of the PDF file**. Second, it should include the source code of your implementation **as the second part of the PDF file** (i.e., portions indicated by “Code”).¹ We will only review the code in detail when the report is not a self-contained document, in order to provide partial credit.

Collaboration versus Academic Misconduct: Collaboration with other students is permitted, but the work you submit must be your own. Copying/plagiarizing work from another student is not permitted and is considered academic misconduct. For more information about University of Colorado Boulder’s Honor Code and academic misconduct, please visit the [course syllabus](#).

¹We require submitting the code as a PDF to avoid many issues that we have observed in the past with being able to access submitted code. These issues have arisen, in part, because we make no programming language requirements. Issues also have arisen from students not providing read permissions for links to their files; e.g., for Google Colab.