

# Homework 5

[Start Assignment](#)

---

**Due** Wednesday by 11:59pm    **Points** 100    **Submitting** a file upload    **Available** Apr 10 at 12am - Apr 20 at 11:59pm

---

## Homework 3 (100 Points)

### Homework instructions:

- You will submit: (1) a report (pdf file) that explains your answers for each problem. The report **MUST** contain the link to your Github repository to access the source code you have developed for this homework. Your report should contain your name, student ID, and homework number. (2) a second pdf file that contains both source codes, results, and plots you have developed and plotted for this homework. You can easily generate this pdf in Jupyter Lab or Jupyter Notebook (sample pdf provided on Canvas.). The pdf file name should contain your name, student ID, and homework number.
- In your report, provide separate and clear responses for each problem. Make reasonable assumptions where necessary and clearly state them! Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.
- You may discuss concepts with your classmates. This fosters group learning and improves the class' progress. However, make sure to submit your own independent and individual solutions.
- **Make sure to use Google Collab (with GPU addon option), or your personal GPU card for the training.**

### Problem 1 (40pts)

For the problem of Machine Translation using sequence-to-sequence model

1- Can you adjust the hyperparameters to improve the translation results?

2. If the encoder and the decoder differ in the number of layers or the number of hidden

units, how can we initialize the hidden state of the decoder? Please run an experiment for it with 3 layers for encoder and 2 layers for decoder. Plot your training results and compare it against the baseline examples from the lectures.

3. Rerun the baseline experiment by replacing GRU with LSTM. Plot your results and compare GRU against LSTM.

### Problem 2 (60pts)

For the problem of Machine Translation with Bahdanau attention based sequence-to-sequence modeling

1. Explore the impacts of number of hidden layers starting from 1 hidden layer up to 4 hidden layers. Plot the results (training loss and validation), also run few examples to do the qualitative comparison between these two. Can you draw the attention weight matrixes and compare them.

2. Replace GRU with LSTM in the experiment. Perform training again. Plot the results (training loss and validation), also run few examples to do the qualitative comparison between these two. Can you draw the attention weight matrixes and compare them.

### Problem 3 (50 Bonus Points)

Can you replace Bahdanau with Luong attention with GRU RNNs, and compare the results against these two while keeping the other dimensions or hyperparameters the same. Report and compare your training results.