

18-848B Lab 2

Lab 2 Report Template

Status: Draft

Authors: Tianshu Huang

Reviewers: Tianshu Huang

Revision History:

Date	Description
Sep 12, 2026	Template created
Sep 15, 2026	Lab posted

Proposed Change

Description

State precisely what you are changing.

Motivation

Argue why this change could improve the model's performance.

Implementation

Describe, at a high level, how you implemented your change: what you added, where it lives in the repository, and how it is selected from the config. Also, describe what unit tests you added for your implementation.

Ablation Study

Results

Split	Baseline test loss	Test loss	Difference	Standard Error of the Difference	Significant (p<0.05)?
10%					
20%					
50%					

Split	Baseline test loss	Test loss	Difference	Standard Error of the Difference	Significant (p<0.05)?
100%					

Scaling Plot

Include your scaling plot here.

Analysis

Q1. Why can't we compute error bars directly using the standard error, $\sqrt{\text{Var}(X) / n}$?

What will happen if we try to do this?

Q2. Why do we need to do a paired comparison?

What would happen if we instead computed the error bars for each method independently, and then compared the two?

Q3. Is your change significantly better?

Is your change better, worse, or statistically indistinguishable from the baseline across the different training set sizes?

Q4. Does your change scale?

Does the gap between the two curves hold, widen, or shrink as the training set grows?

Q5. As a whole, do you believe your change is an improvement?

If so, why? If not, why not?

Q6. Do you think the evaluation set is the right size?

Using your results and/or training curves, argue for whether the evaluation set is too small, too large, or about right.

Q7. Do you think the validation set is the right size?

Using your results and/or training curves, argue for whether 20% is too small, too large, or about right.

