CSCI 145 Problem Set 5

September 17, 2025

Submission Instructions

Please upload *your* work by **11:59pm Monday September 29, 2025.**

- You are encouraged to discuss ideas and work with your classmates. However, you **must acknowledge** your collaborators at the top of each solution on which you collaborated with others and you **must write** your solutions independently.

- Your solutions to theory questions must be written legibly, or typeset in LaTeX or markdown. If you would like to use LaTeX, you can import the source of this document (available from the course webpage) to Overleaf.

- I recommend that you write your solutions to coding questions in a Jupyter notebook using Google Colab.

- You should submit your solutions as a **single PDF** via the assignment on Gradescope.

Grading: The point of the problem set is for *you* to learn. To this end, I hope to disincentivize the use of LLMs by **not** grading your work for correctness. Instead, you will grade your own work by comparing it to my solutions. This self-grade is due the Friday *after* the problem set is due, also on Gradescope.

Problem 1: Regularization

In this problem, we'll investigate the effect of regularization in optimization. Consider an optimization problem with d weights $\mathbf{w} \in \mathbb{R}^d$. Let $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ be a loss function, and $\lambda > 0$ is a regularization hyperparameter. Define the optimal weights

$$\mathbf{w}_\lambda^* = \arg\min_{\mathbf{w} \in \mathbb{R}^d} \mathcal{L}(\mathbf{w}) + \lambda \|\mathbf{w}\|_2^2. \tag{1}$$

Suppose that $\lambda_1 \geq \lambda_2$.

Part A: Regularization Term

Show that

$$\|\mathbf{w}_{\lambda_1}^*\|_2^2 \geq \|\mathbf{w}_{\lambda_2}^*\|_2^2. \tag{2}$$

Hint: Use the optimality conditions i.e., the optimal solution to the optimization problem must be better than every other solution. Plus some algebra.

Part B: Objective Term

Show that

$$\mathcal{L}(\mathbf{w}_{\lambda_1}^*) \leq \mathcal{L}(\mathbf{w}_{\lambda_2}^*). \tag{3}$$

Hint: Use the previous part, plus some algebra.

Part C: Empirical Check

Adapt your code from the logistic regression problem last week so that the loss function is binary cross entropy (with logits) plus a regularization term. Write a function to return the training loss and ℓ_2 norm of the weights for every epoch with hyperparameter λ.

Make one plot with the objective term (BCE) by iteration and one plot with the regularization term by iteration for some $\lambda_1 > \lambda_2$ of your choice. Does the theory we did in the prior parts hold? If not, what went wrong?