CSCI 145 Problem Set 2

September 8, 2025

Submission Instructions

Please upload *your* work by **11:59pm Monday September 15, 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: Gradient Descent on Quadratics

Part A: Closed-form iterates

Consider minimizing a one-dimensional quadratic loss

$$\mathcal{L}(w) = \frac{a}{2}w^2 - b\,w + c \quad \text{with } a > 0.$$

Let $w^* = \arg\min_w \mathcal{L}(w) = b/a$ and the gradient descent update

$$w^{(t+1)} \leftarrow w^{(t)} - \alpha \nabla \mathcal{L}(w^{(t)}).$$

Show that the error contracts geometrically:

$$w^{(t)} - w^* = (1 - \alpha a)^t \,(w^{(0)} - w^*).$$

For which α does this converge? What choice of α gives the fastest (one-step) convergence in 1D?

Part B: Multi-dimensional extension

Let $\mathcal{L}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^\top \mathbf{A}\mathbf{w} - \mathbf{b}^\top \mathbf{w} + c$ with $\mathbf{A} \in \mathbb{R}^{d\times d}$ symmetric positive definite (all eigenvalues are strictly positive). Show that for the gradient descent update $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha(\mathbf{A}\mathbf{w}^{(t)} - \mathbf{b})$,

$$\|\mathbf{w}^{(t)} - \mathbf{w}^*\|_2 \le \rho^t \|\mathbf{w}^{(0)} - \mathbf{w}^*\|_2 \quad \text{where} \quad \rho = \max_i |1 - \alpha\lambda_i(\mathbf{A})|.$$

Deduce that convergence holds iff $0 < \alpha < 2/\lambda_{\max}(A)$.

Hint: Use the inequality that $\|\mathbf{M}\mathbf{v}\|_2 \le \|\mathbf{M}\|_2\|\mathbf{v}\|_2$, for matrix $\mathbf{M}$ and vector $\mathbf{v}$, where $\|\mathbf{M}\|_2 = \lambda_{\max}(\mathbf{M})$.

Part C: Least squares specialization

For mean squared error $\mathcal{L}(\mathbf{w}) = \frac{1}{2n}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$, identify $\mathbf{A}$ and $\mathbf{b}$ and express $\lambda_{\max}(\mathbf{A})$ in terms of $\mathbf{X}$. What step-size bound in terms of $\|\mathbf{X}\|_2$ guarantees convergence?

Part D: Empirical Checks

Load a regression dataset of your choice. With random initializations, repeatedly run gradient descent for 100 epochs with various learning rates. In particular, choose α at evenly spaced intervals from 0 to twice the bound you found in the prior part. Plot the loss by these α. What do you notice?