

CSCI 145 Problem Set 3

September 15, 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

Let's apply gradient descent to quadratic loss functions. Recall the gradient descent update is

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

Part A: Closed-form Iterates

Consider minimizing a one-dimensional quadratic loss $\mathcal{L}(w) = \frac{a}{2}w^2 - bw + c$ with $a > 0$.

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?

Hint: Find w^* , and write b in terms of w^* .

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

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

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

Hint: Follow the same approach as before. Along the way, use the inequality that $\|\mathbf{M}\mathbf{v}\|_2 \leq \|\mathbf{M}\|_2 \|\mathbf{v}\|_2$, for matrix \mathbf{M} and vector \mathbf{v} , where $\|\mathbf{M}\|_2 = \lambda_{\max}(\mathbf{M})$ is the *spectral* norm of a matrix.

Part C: Least Squares Specialization

For the MSE loss $\mathcal{L}(\mathbf{w}) = \frac{1}{2}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$, identify \mathbf{A} and \mathbf{b} and express $\lambda_{\max}(\mathbf{A})$ in terms of \mathbf{X} . What step-size bound in terms of the singular value decomposition of \mathbf{X} 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?

Problem 2: Polynomial Regression and Projections

In this problem, we'll explore polynomial regression, and what could happen if you accidentally delete your labels.

Part A: Polynomial Regression Helps

Load a dataset of your choice. Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be the data matrix, and $\mathbf{y} \in \mathbb{R}^n$ be the target vector. Compute the optimal weights

$$\mathbf{w}_{\text{orig}}^* = \mathbf{X}^+ \mathbf{y}, \quad (2)$$

and output the MSE.

Suppose you want to lower your MSE. As you remember from our class on polynomial regression, you can add new features to your dataset that are derived from higher order functions of the features. Build an augmented data matrix $\mathbf{X}_{\text{aug}} = [\mathbf{X} \quad \mathbf{Z}]$, where $\mathbf{Z} \in \mathbb{R}^{n \times d'}$ contains d' new features of your choice. (You can build these by multiplying various columns of \mathbf{X} together.) Compute the optimal weights

$$\mathbf{w}_{\text{aug}}^* = \mathbf{X}_{\text{aug}}^+ \mathbf{y}, \quad (3)$$

and output the MSE. Ensure that the MSE of the predictions made with the original linear model are higher than the MSE of the predictions made with the augmented linear model.

Part B: Oops

Suppose your labels \mathbf{y} contain sensitive information, and so you delete them after learning $\mathbf{w}_{\text{aug}}^*$. But wait! You actually need the weights $\mathbf{w}_{\text{orig}}^*$ to prove that your new weights $\mathbf{w}_{\text{aug}}^*$ are better. Instead of re-collecting \mathbf{y} , let's try to use the predictions $\mathbf{X}_{\text{aug}} \mathbf{w}_{\text{aug}}^*$ as a stand in for \mathbf{y} . Compute the projected weights

$$\mathbf{w}_{\text{proj}}^* = \mathbf{X}^+ (\mathbf{X}_{\text{aug}} \mathbf{w}_{\text{aug}}^*), \quad (4)$$

and output the MSE. You (should) notice that the MSE of $\mathbf{w}_{\text{proj}}^*$ is *precisely* the same as the MSE of $\mathbf{w}_{\text{orig}}^*$.

How could this be? Compare $\mathbf{w}_{\text{proj}}^*$ to $\mathbf{w}_{\text{orig}}^*$. They're the same!

Part C: SVDs in Action

Remembering the handy SVD we covered in class, you set out to *prove* that $\mathbf{w}_{\text{proj}}^* = \mathbf{w}_{\text{orig}}^*$ always. For simplicity, we'll assume that all the new columns in \mathbf{Z} are orthogonal to the columns in \mathbf{X} i.e., $\mathbf{X}^\top \mathbf{Z} = \mathbf{0}$.

Let $\mathbf{X} = \sum_{i=1}^d \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ be the SVD of \mathbf{X} , and $\mathbf{Z} = \sum_{i=1}^{d'} \sigma'_i \mathbf{u}'_i \mathbf{v}'_i{}^\top$ be the SVD of \mathbf{Z} . Compute $\mathbf{X}_{\text{aug}} \mathbf{X}_{\text{aug}}^\top$ in terms of their SVDs. By our orthogonality assumption, you have just computed the eigendecomposition of $\mathbf{X}_{\text{aug}} \mathbf{X}_{\text{aug}}^\top$.

Hint: Look up how the transpose operation applies to *block* matrices like $\mathbf{X}_{\text{aug}} = [\mathbf{X} \quad \mathbf{Z}]$.

Part D: Projection

Observe that

$$\mathbf{w}_{\text{proj}}^* = \mathbf{X}^+ \mathbf{X}_{\text{aug}} \mathbf{X}_{\text{aug}}^+ \mathbf{y}. \quad (5)$$

Compute $\mathbf{X}_{\text{aug}} \mathbf{X}_{\text{aug}}^+$ in terms of the eigendecomposition of the prior part. Finally, using the SVD of \mathbf{X} , show that $\mathbf{w}_{\text{proj}}^* = \mathbf{X}^+ \mathbf{y}$.