CSCI 145 Problem 13: Can Layer Normalization Destroy All Gradients?

October 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, October 21**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- A complete, good-faith attempt earns submission credit. The grader may select any recent assigned problem for a brief conversation, so keep your work and be prepared to explain each solution and answer a related question. Scheduling details will be announced in class.

Problem 13: Can Layer Normalization Destroy All Gradients?

Layer normalization transforms $\mathbf{z} \in \mathbb{R}^d$ by

$$N(\mathbf{z}) = \frac{\mathbf{z} - \mu\mathbf{1}}{\sigma}, \qquad \mu = \frac{1}{d}\sum_{i=1}^{d} z_i, \qquad \sigma^2 = \frac{1}{d}\sum_{i=1}^{d}(z_i - \mu)^2.$$

Here $\mathbf{1}$ is the vector of ones. Assume $\sigma > 0$ and omit the small constant usually added to the variance for numerical stability.

1. Suppose $d = 2$, so $\mathbf{z} = (a, b)$ with $a \neq b$. Find $N(\mathbf{z})$.
2. Show that the Jacobian of N is the zero matrix whenever $d = 2$ and $a \neq b$. What does this mean for gradients flowing through the layer?
3. For general $d \geq 2$, show that the Jacobian of N has rank at most $d - 2$.

Hint: For part 3, what happens when you add the same constant to every coordinate, or multiply every coordinate by a positive constant?

Takeaway: *Normalization removes changes in the common offset and scale. With just two coordinates, it keeps only their ordering, so small input changes have no effect and no gradient passes through. In higher dimensions, these same two directions are still lost.*