CSCI 145 Problem 11: Expressivity versus Optimization

August 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on DATE**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- Problems are grouped into three-week **cycles**. Once per cycle, you will meet with the grader and present the solution to *one* problem from that cycle, selected at random when you arrive; bring your work and be ready to explain any of them.

Problem 11: Expressivity versus Optimization 2 points

Consider the XOR data set $\mathbf{x} \in \{0,1\}^2$, with label $y = 1$ when exactly one coordinate of $\mathbf{x}$ is one and $y = -1$ otherwise. Fit it with

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = c + \sum_{j=1}^{2} a_j \, \mathrm{ReLU}(\mathbf{w}_j^\top \mathbf{x} + b_j)$$

by minimizing the average squared error over the four examples.

A researcher claims: "Once a neural-network architecture can fit the training data, every parameter setting with nonzero loss has a descent direction." Determine whether this architecture can fit the data, then decide whether the claim is true for its squared-error landscape. Give a proof or counterexample that accounts for the full gradient with respect to c, a_j, $\mathbf{w}_j$, and b_j. Any stationary point used in your argument must keep every hidden preactivation strictly away from the ReLU kink on all four inputs, so nondifferentiability cannot explain it.

Your answer should make clear which fact is about the functions the architecture can represent and which is about the loss landscape seen by backpropagation.

Results you may use. On a strictly negative input, the ReLU and its derivative are both zero. You may also quote the scalar backpropagation equations from the reading.