CSCI 145 Problem 11: Depth Folds Space

July 20, 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 — so bring your work and be ready to explain any of them.

Problem 11: Depth Folds Space

Lecture's XOR network folded the input plane into a tent with two hidden neurons. Now suppose you are handed twenty ReLU neurons, to arrange either as one hidden layer of twenty or as ten layers of two. This problem counts the straight pieces each arrangement can build, and the two answers are not close.

Throughout, the input is one-dimensional, $x \in \mathbb{R}$, and every network is built from ReLU units and linear combinations. Following the reading's notation, hidden unit i computes $h_i(x) = \mathrm{ReLU}(w_i^{(1)}x + b_i^{(1)})$ with $w_i^{(1)}, b_i^{(1)} \in \mathbb{R}$, and a network with one hidden layer of width m outputs

$$f(x) = \sum_{i=1}^{m} w_i^{(2)} h_i(x) + b^{(2)}.$$

Call a continuous function *piecewise linear with k pieces* if the line splits into k intervals on each of which it is a straight line, with the $k-1$ interior boundaries as its *kinks*.

Part A: One Layer, m Neurons

A single ReLU unit h_i is continuous and piecewise linear with at most one kink (none at all if $w_i^{(1)} = 0$), at the point $x = -b_i^{(1)}/w_i^{(1)}$ where it switches from flat to sloped. Explain why f is then piecewise linear with at most $m+1$ pieces. Then argue that f can be built to pass through any $m+1$ points $(x_1, y_1), \ldots, (x_{m+1}, y_{m+1})$ with distinct, increasing x_i. Comparing the kinks available against the segments an interpolation needs is enough; you need not solve for the parameters. This is a constructive universal approximation statement: enough width hits any finite set of points exactly.

Part B: The Sawtooth

Now hold the width at two neurons per layer and spend the budget on depth instead, starting from the tent function

$$g(x) = 2\,\mathrm{ReLU}(x) - 4\,\mathrm{ReLU}\big(x - \tfrac{1}{2}\big),$$

two hidden units plus a linear output, and exactly the lecture's XOR network read in the variable $x = (x_1 + x_2)/2$. Verify $g(0) = 0$, $g(\frac{1}{2}) = 1$, and $g(1) = 0$, so g maps $[0,1]$ onto $[0,1]$.

Let $g^{(L)} = g \circ g \circ \cdots \circ g$ be L copies of g stacked as L separate layers, a network of $2L$ neurons in total. Prove by induction on L that $g^{(L)}$ has 2^L linear pieces on $[0,1]$. The base case $L = 1$ has two pieces, up and then down; for the inductive step, on each piece of $g^{(L-1)}$ the output sweeps monotonically from 0 to 1 or from 1 to 0, so composing with g folds that piece into two. Finish by evaluating $L = 10$: how many pieces, out of how many neurons?

Part C: What the Same Shape Costs in Width

Part A says each hidden unit contributes at most one kink, so run that count in reverse. Show that any one-hidden-layer network agreeing with $g^{(L)}$ on $[0,1]$ needs at least $2^L - 1$ hidden units, and compare that at $L = 10$ against the $2L$ neurons Part B used. Conclude the punchline in one sentence: at a fixed neuron budget, what does depth buy that width cannot?

Where this goes next. The composition that multiplies pieces is the same mechanism that makes gradients vanish or explode through many layers, which is the problem the Depth-enablers lecture spends a session on.