Final exam. Problem 1.

This problem requires both analytical calculations and numerical experiments. Submit a pdf file with your report and figures and link your codes to it. All your claims regarding convergence should be supported with numerical evidence.

Suppose we want to approximate to function $g(x) = 1 - \cos x$ on the interval $[0, \pi/2]$ with the function $\mathsf{ReLU}(ax - b)$ where a and b are to be determined. We take 6 training points $x_j = \pi j/10$, $j = 0, 1, 2, 3, 4, 5$, and set up the following loss function:

$$f(a, b) = \frac{1}{12} \sum_{j=0}^{5} \left[\mathsf{ReLU}(ax_j - b) - g(x_j)\right]^2. \tag{1}$$

1. The set of stationary points of f, i.e., the set of points where $\nabla f = 0$, consists of the global minimizer and a flat region. Describe this set analytically using equalities and inequalities and show it in a figure. Provide an analytic formula for the global minimizer of f. What is the global minimum of f?
2. Take $a = 1$ and $b = 0$ as the initial guess for the gradient descend with constant stepsize. What is the minimal stepsize α^* such that the iterates end up in the flat region? Suppose we take $\alpha = 0.99\alpha^*$ and run the gradient descend. Will the iterates approach the global minimizer? Either way, explain why.

   Propose a stepsize trying to make it as large as possible, and give a rationale for your choice, such that the iterates will necessarily converge to the global minimizer.
3. As above, take $a = 1$ and $b = 0$ as the initial guess. Use a simple stochastic gradient descend with a single training point chosen randomly for approximating the gradient of f at each step. Find a strategy for stepsize reduction such that the stochastic gradient descend will converge to the global minimizer.