Algebraic Approach to Ridge-Regularized Mean Squared Error Minimization in Minimal ReLU Neural Network

Akifumi Okuno (ISM, SOKENDAI, RIKEN)

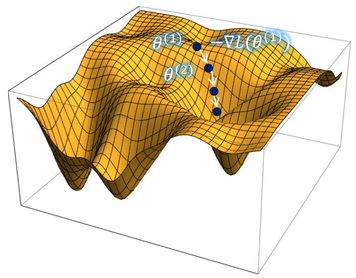
Joint work with Ryoya Fukasaku (Kyushu U.) and Yutaro Kabata (Kagoshima U.)

1. Problem Setting

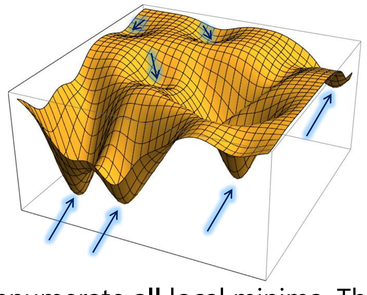
$\mathrm{ReLU}(z) = \max\{0, z\}$

Consider training a simple ReLU perceptron $f_\theta(x) = [\![a, \mathrm{ReLU}(Bx_i + c)]\!]$ by minimizing ridge-regularized loss:

$$\tilde{\ell}_\lambda(\theta) = \sum_{i=1}^{n} \big(y_i - f_\theta(x_i)\big)^2 + \lambda \|\theta\|_2^2.$$



Typically, the loss function, which is highly non-convex and rugged, is optimized by gradient-based methods, yielding one local minimum as the outcome.



【Our goal】 is to enumerate **all** local minima. The minima are not necessarily isolated (i.e., nonzero-dimensional solution sets are allowed).

2. Our Strategy

With the activation pattern $e_{i\ell} = e_{i\ell}(\theta, x_i) = \mathbb{I}(\langle b_\ell, x_i \rangle + c_\ell \geq 0)$, ReLU-activated function reduces to $\mathrm{ReLU}(\langle b_\ell, x_i \rangle + c_\ell) = e_{i\ell}(\langle b_\ell, x_i \rangle + c_\ell)$. Namely, if the activation pattern $E = (e_{i\ell}) \in \{0,1\}^{n \times L}$ is given, the perceptron becomes a polynomial function, which implies that the loss function $\tilde{\ell}_{\lambda,E}(\theta)$ is also polynomial (called **algebraic surrogate**). Therefore, any local minimum should satisfy the polynomial equation:

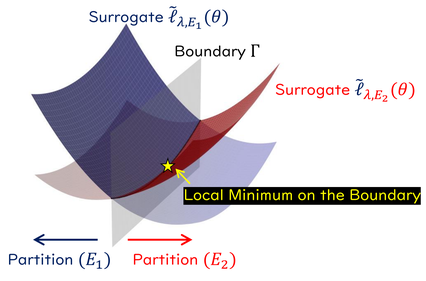
The derivative is also a polynomial.

$$\frac{d\tilde{\ell}_{\lambda,E}(\theta)}{d\theta} = 0$$

All solutions of the polynomial equations can be identified using computational algebra (especially, Groebner basis).

However, the solution does not necessarily correspond to the assumed activation pattern E.
In our approach, we first list all possible activation patterns $E_1, E_2, \ldots, E_{nL}$, solve the corresponding polynomial equations, and then filter out solutions that do not satisfy the respective activation patterns.
Within the interior of the parameter region corresponding to each activation pattern, **this method can enumerate all local minima candidates.**

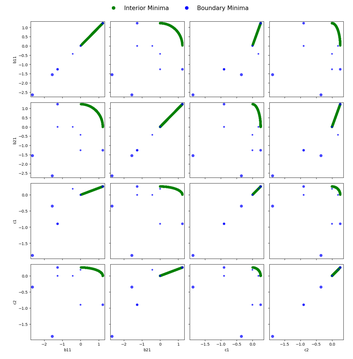
3. Remaining Problem: Boundary Minima



The algebraic surrogate coincides with the original loss function within the parameter region corresponding to each activation pattern. However, at the boundaries where the surrogate switches between regions, the function can change abruptly and non-smoothly, potentially giving rise to local solutions.
Such **boundary local solutions can also be enumerated using Lagrange multipliers**, which are themselves closed under polynomial form, allowing **these solutions to be exhaustively enumerated through computational algebra**.

4. Numerical Demonstration

We computed the solutions for regression using a ReLU neural network with $L = 2$ units and a dataset of size $n = 5$. After partitioning the parameter space into $2^{nL} = 1024$ activation patterns, we obtained several local minima. Among the solutions, **eight were zero-dimensional and all located on the boundaries between activation regions**. In contrast, **a single one-dimensional continuous solution appeared within the interior** of one activation region.



Blue indicates eight minima located on the boundaries, while green represents a one-dimensional minimum within the interior.

For more details, see
https://arxiv.org/abs/2508.17783