

Basic idea to reformulate Rectified Linear Unit (ReLU)

$$\text{ReLU}(x) = \ln(1 + \exp(x))$$

In an MLP, from one layer to the next, each node activation $\mathbf{z} = \mathbf{W}\mathbf{h} + \mathbf{b}$
WLOG, let the bias $\mathbf{b} = 0$. Also, assume ReLU is applied elementwise to the vector \mathbf{z}

Then, for each activated node \mathbf{n}_i

$$n_i = \text{ReLU}(\mathbf{z}_j) = \ln(1 + \exp((\mathbf{W}\mathbf{h})_j))$$

$$= \ln(1 + \exp((\sum_i \mathbf{W}_{ij}\mathbf{h}_i)))$$

$$= \ln(1 + \prod_i \exp((\mathbf{W}_{ij}\mathbf{h}_i)))$$

Which seems intractable

But

What if we consider $\text{ReLU}(\mathbf{z}_j)$ in Expectation

$$\mathbb{E}[\text{ReLU}(\mathbf{z}_j)]$$

we assume that each activated node exists in a superposition of 2 states $[0,1]$
what I want to do is somehow get the superposition inside the log, so that

$$\mathbb{E}[\text{ReLU}(\mathbf{z}_j)] \approx \ln(\mathbb{E}[1 + \exp((\mathbf{W}\mathbf{h})_j)])$$

this is kind of like the annealed approximation in spin glass theory—which is not a great approximation, but perhaps, can give some initial insight
from here, let's introduce an internal state vector $\mathbf{c} = [0, 1]$

$$\ln(\mathbb{E}[1 + \exp((\mathbf{W}\mathbf{h})_j)]) = \ln(\mathbb{E}[\exp((\mathbf{c}^T \mathbf{W}\mathbf{h})_j)])$$

and we can get something like

$$\mathbb{E}[\text{ReLU}(\mathbf{z}_j)] \approx \ln(\mathbb{E}[\exp((\mathbf{c}^T \mathbf{W}\mathbf{h})_j)])$$

$$\approx \mathbb{E}[\ln(\exp((\mathbf{c}^T \mathbf{W}\mathbf{h})_j))]$$

$$\approx \mathbb{E}[(\mathbf{c}^T \mathbf{W}\mathbf{h})_j]$$

So obviously this is very sloppy, but it is an attempt to get at defining some kind of (linearized) free energy for each layer