Basic idea to reformulate Recitfied Linear Unit (ReLu)

$$ReLu(x) = ln(1 + exp(x))$$

In an MLP, from one layer to the next, each node activation $\mathbf{z} = \mathbf{Wh} + \mathbf{b}$

WLOG, let the bias $\mathbf{b} = 0$. Also, assume ReLu is applied elementwise to the vector $\mathbf{z}$

Then, for each activated node $\mathbf{z}_i$

$$ReLu(\mathbf{z}_j) = ln(1 + exp((\mathbf{Wh})_j)$$

$$ReLu(\mathbf{z}_j) = ln(1 + exp((\sum_i \mathbf{W_{ij}h_i}))$$

$$ReLu(\mathbf{z}_j) = ln(1 + \prod_i exp((\mathbf{W_{ij}h_i}))$$

Which seems intractable

But

What if want consider $Relu(\mathbf{z}_j)$ in Expectation

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

and

we assume that each activated node $\mathbf{z}_i$ exists in a superposition of 2 states [0,1]