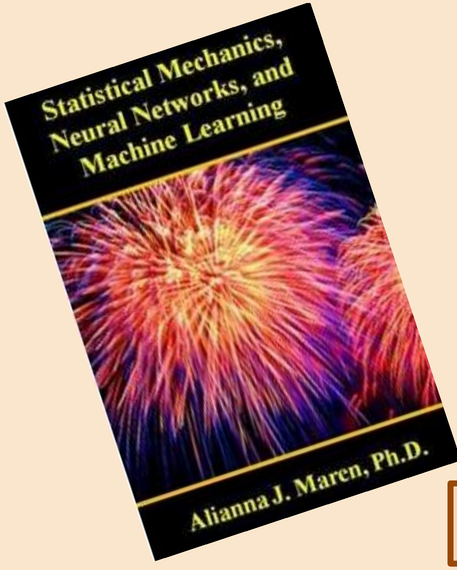
Statistical Mechanics for Neural Networks and Deep Learning

with

Alianna J. Maren, Ph.D.

From Microstates to Partition Functions: *Including Examples*



Author's Note:

The following material is ***yet one more accompaniment*** for the *Précis* for the book-in-progress, **Statistical Mechanics for Neural Networks and Deep Learning**.
For more information, Opt-In at:
www.aliannajmaren.com

Alianna J. Maren, Ph.D.

Our Goal:

Compute the probabilities for each of the microstates, and then the overall likelihood for the system to be in a given configuration.

From there, identify which configuration is the most likely one for the system when we have specified a value for the upper level energy, e_j ***.***

To do this, we first need the partition function.

Recall our Probability and Partition Function Equations:

Probability:

$$p_j = \frac{1}{Z}\exp(-\beta E_j)$$

Partition Function:

$$Z = \sum_j \exp(-\beta E_j)$$

NOTE: Because the probability function uses the partition function (Z) as a normalizing factor, we have to compute the partition function first.

Thus, our first step is to obtain $\exp(-\beta E_j)$ ***for each configuration (number of units in a given energy level) for our system.***

Recall that E_j ***is the total energy for the system in that state. That means,*** $E_j = M * e_j$ ***, where M is the total number of units in the upper energy level*** e_j ***.***

As always, we let $\beta = 1$ ***.***

Recall Our Microstates Table (from previous slidedeck):

Given a specific energy for the upper level, e_1, we can compute the energy for that configuration.

System state	n_0 = # Units at e_0	n_1 = # Units at e_1	Total number of microstates
1	10	0	1
2	9	1	10
3	8	2	45
4	7	3	120
5	6	4	210
6	5	5	**252**
7	4	6	210
8	3	7	120
9	2	8	45
10	1	9	10
11	0	10	1

Symmetric Results,
we get the same results when we have a certain number of units on the upper level as we when we have those same units on the lower level.

Maximal number of microstates when equal numbers in upper and lower energy levels.

This will connect with our notion of entropy later on.

Computing Actual Probabilities for Each Configuration:

Divide through by Z, which we just computed; ***Z = 21.944***.

System state	*M* = # Units at e_1	*Q* = Total number of microstates	$E_j = M * e_1$ Use: $e_1 = 1$	$\exp(-\beta E_j)$ $\beta = 1$	$Q * \exp(-\beta E_j)$	p_j	Total likelihood = $Q*p_j$
1	0	1	0	1.000	1.000	0.0456	0.046
2	1	10	1	0.368	3.680	0.0168	0.168
3	2	45	2	0.135	6.075	0.0062	0.279
4	3	120	3	0.050	6.000	0.0023	0.276
5	4	210	4	0.018	3.780	0.0008	0.168
6	5	**252**	5	**0.007**	**1.764**	**--**	**--**
7	6	210	6	0.0025	0.525	--	**--**
8	7	120	7	0.001	0.120	--	**--**
9	8	45	8	--	--	**--**	**--**
10	9	10	9	--	--	**--**	**--**
11	10	1	10	--	--	**--**	**--**

Interpretation: Winning energy level is for *three* units to be in energy level $e_1 = 1$, followed closely by *four* units at that level.

Computing Likelihoods for Each Configuration:

Given a specific energy for the upper level, e_1, we can compute the energy for that configuration.

System state	M = # Units at e_1	Q = Total number of microstates	$E_j = M * e_1$ $e_1 = 1$	$\exp(-\beta E_j)$ $\beta = 1$	$Q * \exp(-\beta E_j)$
1	0	1	0	1.000	1.000
2	1	10	1	0.368	3.680
3	2	45	2	0.135	6.075
4	3	120	3	0.050	6.000
5	4	210	4	0.018	3.780
6	5	**252**	5	**0.007**	**1.764**
7	6	210	6	0.0025	0.525
8	7	120	7	0.001	0.120
9	8	45	8	*Negligible*	*Negligible*
10	9	10	9	*Negligible*	*Negligible*
11	10	1	10	*Negligible*	*Negligible*

Energy level with most microstates is not the one that has the lowest total associated energy!

$$Z = \sum_j \exp(-\beta E_j) = 1.000 + 3.680 + 6.075 + 6.000 + 3.780 + 1.764 + 0.525 + 0.120 = 21.944$$

Interpretation of Results:

We just found the configuration (number of units at an the upper energy level) that was most probable, given a specific value for that energy level e_1:

1. ***It <u>was NOT</u> the level that had the lowest overall energy.***
2. ***It <u>was NOT</u> the level that had the most microstates.***
3. ***<u>It was in-between</u>. The "winning level" was a combination of low total energy AND lots of microstates.***

When we get to free energy minimization, we'll be able to find this energy level much faster – single-shot equation.

What We've Accomplished So Far:

We now know how to do two steps in our journey through the "Rocky Mountains" of energy-based machine learning:

1. ***Compute the partition function** for the system, and then*
2. ***Compute the most likely probability** for a given configuration (i.e., given specific M and (N-M)),*

Based on the probabilities and number of microstates per energy level, we are now able to:

Figure out the most likely system configuration.

This means that we are EXTREMELY CLOSE, in practice, to free energy minimization.

What We've Done So Far:

1. *Partition Function*
2. *Probabilities (statistical mechanics method)*

Next Steps:

3. *Entropy*
4. *Free Energy (& minimization thereof)*
5. *Bayesian probabilities*
6. *Kullback-Leibler Divergence*
7. *Inference (approximation methods, e.g. variational Bayes)*