Thought Curvature: A hypothesis on the 'Supersymmetric Artificial Neural Network'

Jordan Micah Bennett

Abstract

The "Supersymmetric Artificial Neural Network" hypothesis(with the novel-denotation $\phi(x;\theta,\overline{\theta})^{\top}w$) seeks to explore a regime with the potential to fundamentally use supersymmetric methods to construct artificial neural networks, therein seeking to engender novel, non-trivial contributions to deep or hierarchical artificial learning. Looking at the progression of 'solution geometries'; going from $SO(n)$ representation (such as Perceptron like models) to $SU(n)$ representation (such as UnitaryRNNs) has guaranteed richer and richer representations in weight space of the artificial neural network, and hence better and better hypotheses were generatable. The Supersymmetric Artificial Neural Network hypothesis explores a natural step forward, namely $SU(m|n)$ representation. These supersymmetric biological brain representations (Perez et al.) can be represented by supercharge compatible special unitary notation $SU(m|n)$, or $\phi(x;\theta,\overline{\theta})^{\top}w$ parameterized by $\theta,\overline{\theta}$, which are supersymmetric directions, unlike θ seen in the typical non-supersymmetric deep learning model. Notably, Supersymmetric values can encode or represent more information than the typical deep learning model, in terms of "partner potential" signals for example. This paper does not contain empirical code concerning supersymmetric artificial neural networks, although it does highlight empirical evidence, that indicates how such types of supersymmetric learning models could exceed the state of the art, due to preservation features seen in progressing through earlier related models from the days of older perceptron like models that were not supersymmetric.

Keywords. Supersymmetric-gradient-descent, Supermathematics, Supermanifold, Supersymmetry, Lie Superalgebra, Deep Learning

Note. "Thought Curvature" is similar to "Thought vectors[19]", with the distinction that supermanifold[1]/curvatures[21] are used to describe the "Supersymmetric Artificial Neural Network"(SANN) hypothesis. (See manifold[2]/curvature[21] work in geometric deep learning by Michael Bronstein et al.[20])

Introduction. Machine learning non-trivially concerns the application of families of functions that guarantee more and more variations in weight space.

This means that machine learning researchers study what functions are best to transform the weights of the artificial neural network, such that the weights learn to represent good values for which correct hypotheses or guesses can be produced by the artificial neural network.

The "Supersymmetric Artificial Neural Network" (or 'thought curvature') is reasonably yet another way to represent richer values in the weights of the model; because supersymmetric values can allow for more information to be captured about the input space. For example, supersymmetric systems can capture potential-partner signals, which is beyond the feature space of magnitude and phase signals learnt in typical real valued neural nets and deep complex neural networks respectively. As such, a brief historical progression of geometric solution spaces for varying neural network architectures follows:

1. An optimal weight space produced by shallow or low dimension integer valued nodes or real valued artificial neural nets, may have good weights that lie for example, in one simple ($\mathbb{Z}^n$ or $\mathbb{R}^n - ordered$) cone per class/target group. (This may guarantee some variation, but not enough for more sophisticated tasks of higher dimension)[5][10]

2. An optimal weight space produced by deep and high-dimension-absorbing real valued artificial neural nets, may have good weights that lie in disentangleable ($\mathbb{R}^n * \mathbb{R}^n - ordered$) manifolds per class/target group convolved by the operator $*$, instead of the simpler regions per class/target group seen in item (1).

Paper publication date: May 23, 2016. (Updated afterwards)

(This may guarantee more variation in the weight space than (1), leading to better hypotheses or guesses)[6]

3. An optimal weight space produced by shallow but high dimension-absorbing complex valued artificial neural nets, may have good weights that lie in multiple $(\mathbb{C}^n - ordered)$ sectors per class/target group, instead of the real regions per class/target group seen amongst the prior items. (This may guarantee more variation of the weight space than the previous items, by learning additional features, in the "phase space". This also leads to better hypotheses/guesses)[7]

4. An optimal weight space produced by deep or high dimension-absorbing complex valued artificial neural nets, may have good weights that lie in chi distribution bound, $(\mathbb{C}^n * \mathbb{C}^n - ordered)$ rayleigh space per class/target group convolved by the operator $*$, instead of the simpler sectors/regions per class/target group seen amongst the previous items. (This may guarantee more variation of the weight space than the prior items, by learning phase space representations, and by extension, strengthen these representations via convolutional residual blocks. This also leads to better hypotheses/guesses)[8]

5. The "Supersymmetric Artificial Neural Network" operable on high dimensional data, may reasonably generate good weights that lie in disentangleable $(C^\infty(R^{m|n}) - ordered)$ supermanifolds per class/target group, instead of the solution geometries seen in the prior items above. Supersymmetric values can encode rich partner-potential delimited features beyond the phase space of (4) in accordance with cognitive biological space[3], where (4) lacks the partner potential formulation describable in Supersymmetric embedding.[9]

Another view of "solution geometry" history, which may promote a clear way to view the reasoning behind the subsequent pseudocode sequence.

1. There has been a clear progression of "solution geometries", ranging from those of the ancient Perceptron[5] to complex valued neural nets [8], grassmann manifold artificial neural networks[13] or unitaryRNNs.[12][14][16] These models may be denoted by $\phi(x;\theta)^\top w$ parameterized by θ, expressible as geometrical groups ranging from orthogonal[4] to special unitary group[18] based: $SO(n)$ to $SU(n)$..., and they got better at representing input data i.e. representing richer weights, thus the learning models generated better hypotheses or guesses.
2. By "solution geometry" I mean simply the class of regions where an algorithm's weights may lie, when generating those weights to do some task.
3. As such, if one follows cognitive science, one would know that biological brains may be measured in terms of supersymmetric operations. (Perez et al, "Supersymmetry at brain scale" [3])
4. These supersymmetric biological brain representations can be represented by supercharge[11] compatible special unitary notation $SU(m|n)$, or $\phi(x;\theta,\overline{\theta})^\top w$ parameterized by $\theta,\overline{\theta}$ [9], which are supersymmetric directions, unlike θ seen in item (1). Notably, Supersymmetric values can encode or represent more information than the prior classes seen in (1), in terms of "partner potential" signals for example.
5. So, state of the art machine learning work forming $U(n)$ or $SU(n)$ based solution geometries, although *non-supersymmetric*, are already in the family of *supersymmetric* solution geometries that may be observed as occurring in biological brain or $SU(m|n)$ supergroup[17] representation.

A naive supersymmetric artificial neural network architecture. (See points 1 to 5 in [15]) It seems feasible that a C^∞ bound atlas-based learning model, where said C^∞ is in the family of supermanifolds from supersymmetry, may be obtained from a system, which includes charts $\binom{n}{k}$ of *grassmann manifold networks* $GR_{k,n}$ and *stiefel manifolds* $GF_{k,n}$, in (ϕ_I, U_I) terms, where there exists some invertible submatrix entailing

matrix $A \in \phi_I(U_I \cap U_J)$ for $U_I = \pi(V_I)$ where π is a submersion mapping on some *stiefel manifold* $GF_{k,n}$, thereafter enabling some differentiable *grassmann manifold* $GR_k(\mathbb{R}^n)$ wherein $V_I = \{u \in \mathbb{R}^{n\times k}: det(u_I) \neq 0\}$.

Conclusion. Pertinently, the “Edward Witten/String theory powered supersymmetric artificial neural network”, is one wherein supersymmetric weights are sought.

Many machine learning algorithms are not empirically shown to be exactly biologically plausible, i.e. Deep Neural Network algorithms, have not been observed to occur in the brain, but regardless, such algorithms work in practice in machine learning.

Likewise, regardless of Supersymmetry's elusiveness at the LHC, as seen above, it may be quite feasible to borrow formal methods from strategies in physics even if such strategies are yet to show related physical phenomena to exist; thus it may be pertinent/feasible to try to construct a model that learns supersymmetric weights, as I proposed throughout this paper, following the progression of solution geometries going from $SO(n)$ to $SU(n)$ and onwards to $SU(m|n)$ [17].

Appendix: Artificial neural network/symmetry group landscape visualization.

Thought Curvature Figure 1: Let φ denote what each type of model can learn. (See "Deep Learning Book" by Bengio et al. Chapter 6, page 166, point 3 for φ description)

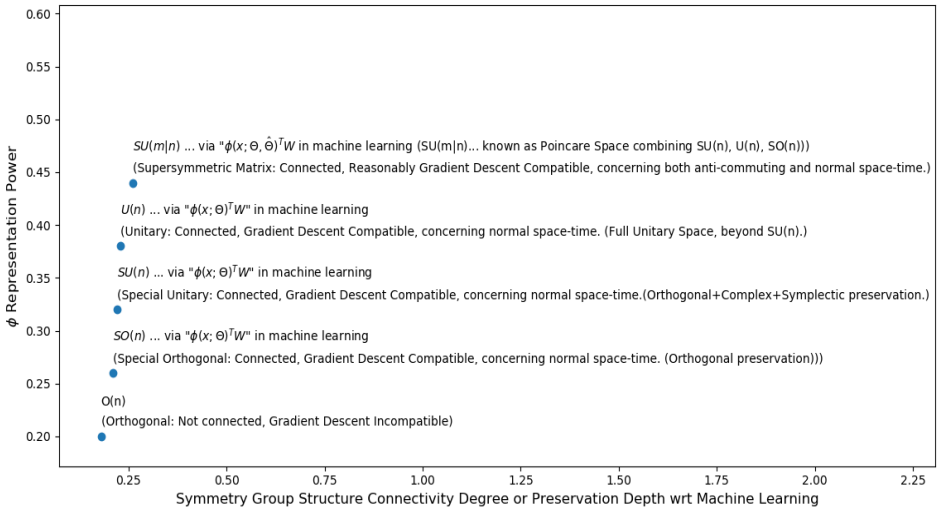

Figure Description: The two axes above are in arbitrary units, that seek to display non-arbitrary progression in representation power and symmetry group connectivity degree. A larger picture is available here.

1. $\boldsymbol{O(n)}$ structure – Orthogonal is not connected enough, therefore *not amenable to gradient descent* in machine learning. (Paper: See note 2 at end of page 2, in reference [22].)
2. $\boldsymbol{SO(n)}$ structure – Special Orthogonal; is connected, *gradient descent compatible*, while preserving orthogonality, concerning *normal space-time*. (Paper: See paper in item 1).
3. $\boldsymbol{SU(n)}$ structure – Special Unitary; is connected, gradient descent compatible; complex generalization of $\boldsymbol{O(n)}$, but only a subspace of larger unitary space, concerning *normal space-time*. (The *Unitary Evolution Recurrent Neural Network*[14] related to *complex unit circle* seen in $U(1)$ in physics (See page 2 in [24].))
4. $U(n)$ structure – Unitary[16]; is connected, *gradient descent compatible*; Larger unitary landscape than $SU(n)$, concerning *normal space-time*.
5. $SU(m|n)$ structure – Supersymmetric; *is connected* [23], thereafter reasonably *gradient descent compatible* and even larger than the $U(n)$ landscape, to permit sparticle invariance, *being a Poincare group extension (See page 7 in* [24]) containing *both normal space-time and anti-commuting components*, as seen in the Supersymmetric Artificial Neural Network which this paper proposes.

References

1. Wikipedia, “Supermanifolds”.
2. Wikipedia, “Differentiable Manifolds”.
3. P´erez et al. “Supersymmetric methods in the travelling variable: inside neurons and at the brain scale”, 2007.
4. Wikipedia “Orthogonal Groups”.
5. Wikipedia “Perceptron”
6. Wikipedia “Deep Learning#Deep Neural Networks”
7. Tariq Rashid “Complex Valued Neural Networks - Experiments”, 2016
8. Chiheb Trabelsi, Olexa Bilaniuk et al. “Deep Complex Networks”, 2017
9. Wikipedia “Supersymmetry”
10. Wikipedia “Artificial Neural Network#Hebbian learning”
11. Wikipedia “Supercharge”
12. Simone Fiori “A study on neural learning on manifold foliations: the case of the Lie group SU(3)”, 2008
13. Zhiwu Huang, Jiqing Wu et. al. “Building Deep Networks on Grassmann Manifolds”, 2017
14. Martin Arjovsky, Amar Shah et. al. “Unitary Evolution Recurrent Neural Networks”, 2015
15. Math wisc edu “The Grassmann Manifold”
16. Li Jing, Yichen Shen, “Tunable Efficient Unitary Neural Networks (EUNN) and their application to RNNs”, 2017
17. Wikipedia “Supergroup_(physics)”
18. Wikipedia “Special unitary group”
19. Wikipedia “Thought Vector
20. Michael Bronstein et. Al. “Geometric deep learning: going beyond Euclidean data”, 2016
21. Wikipedia, “Curvature”
22. Mario Lezcano-Casado et al, “Cheap Orthogonal Constraints in Neural Networks: A Simple Parametrization of the Orthogonal and Unitary Group”, 2019
23. Wikipedia “Supersymmetry algebra”
24. Physics Wisc Edu, “Symmetry Groups”