Feature Extraction of Neural Networks Applied to Magnetic Models

Santiago Salazar Jaramillo

Supervisor: Alonso Botero Mejía, Ph.D.

Departamento de Física, Univ. de Los Andes, Bogotá, Colombia.

s.salazar12@uniandes.edu.co

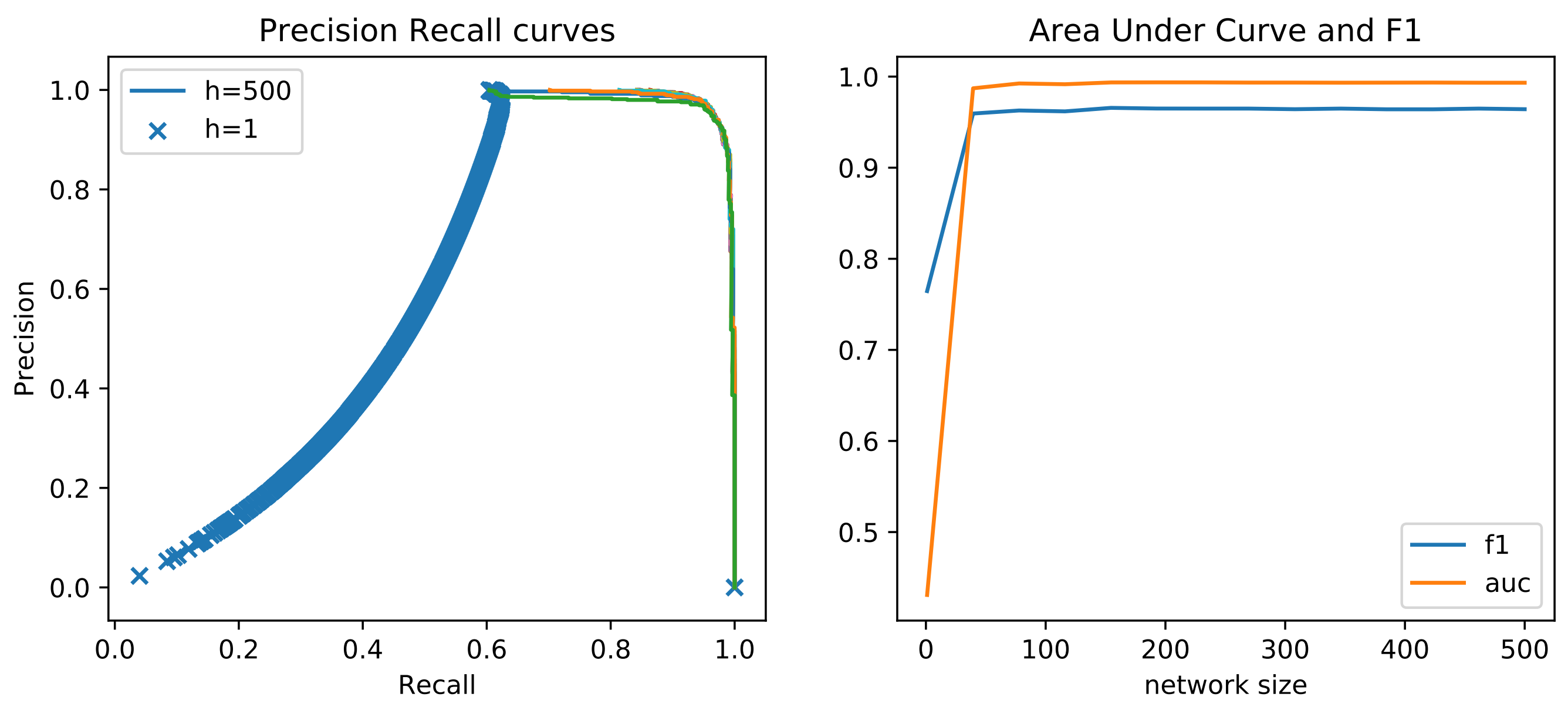
Abstract

Neural Networks have been proven to be highly efficient and versatile algorithms. They are capable of solving tasks by recognising statistical features in the input data, but due to their complexity, NN's are usually treated as a black-box. *Thus, most applications focus solely on the performance of NN's, while ignoring their feature extraction capabilities. In the case of physics, NN's trained to recognise phase transitions of different models, have been shown to recognise order parameters such as magnetisation [4, 2, 1]. In this work, a neural network was trained on spin configurations and it's weight matrices were analysed using other machine learning algorithms, in order to identify which statistical features the algorithm was capable of learning.*

1. Data and Network Architecture

- The input data was made of a set of spin configurations generated by a Montecarlo simulation of the square lattice Ising model [3].
- Each sample was labeled by a 2 component vector according to it's temperature.
- The network was chosen as a densely connected NN, with a L2 regularised cross entropy loss function and Adam optimiser learning rule.
- The hidden layer consisted of 70 neurons.
- The output layer consisted of two neurons, one for each phase, which mirrors the label vector.

2. Performance and Hidden Layer Size



Maximum $F1$ was 0.965675 for 50 hidden neurons. Maximum AUC was 0.993660 for 100 hidden neurons. A sharp increase in performance was found after 3 hidden neurons.

3. Results

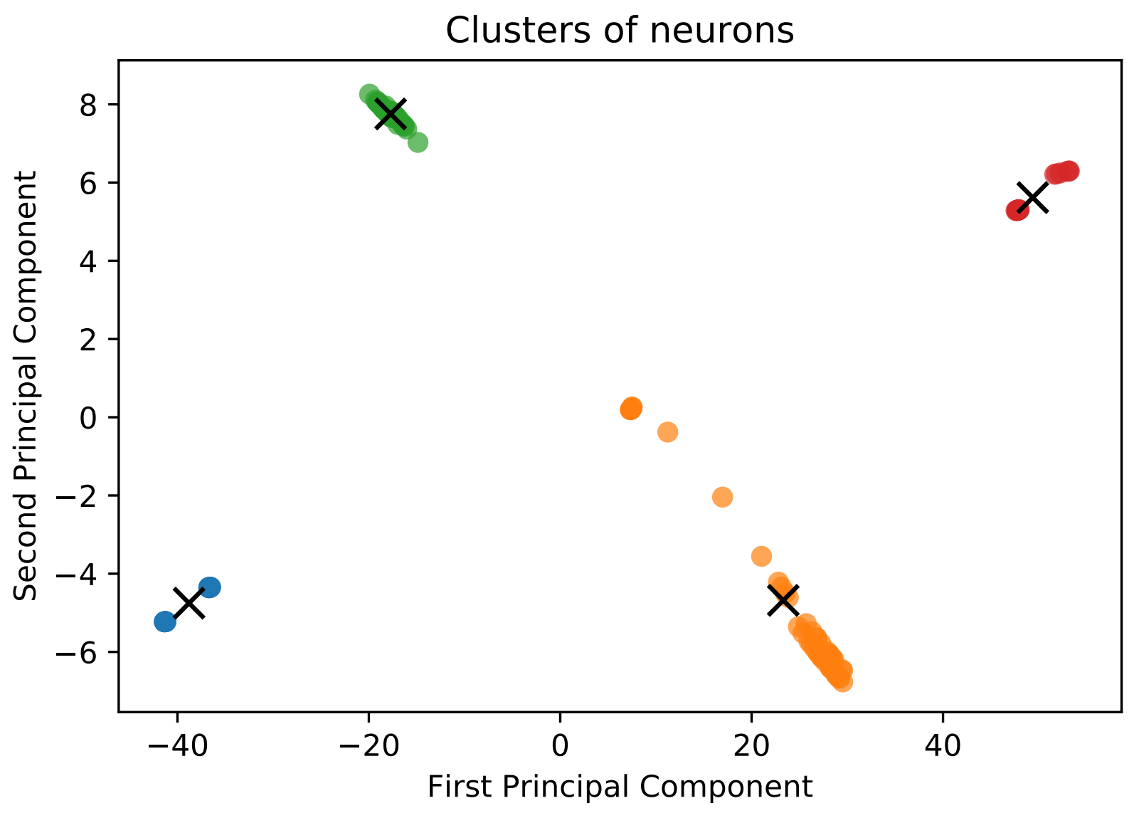


Figure 3: K-means clustering of the weight matrices.

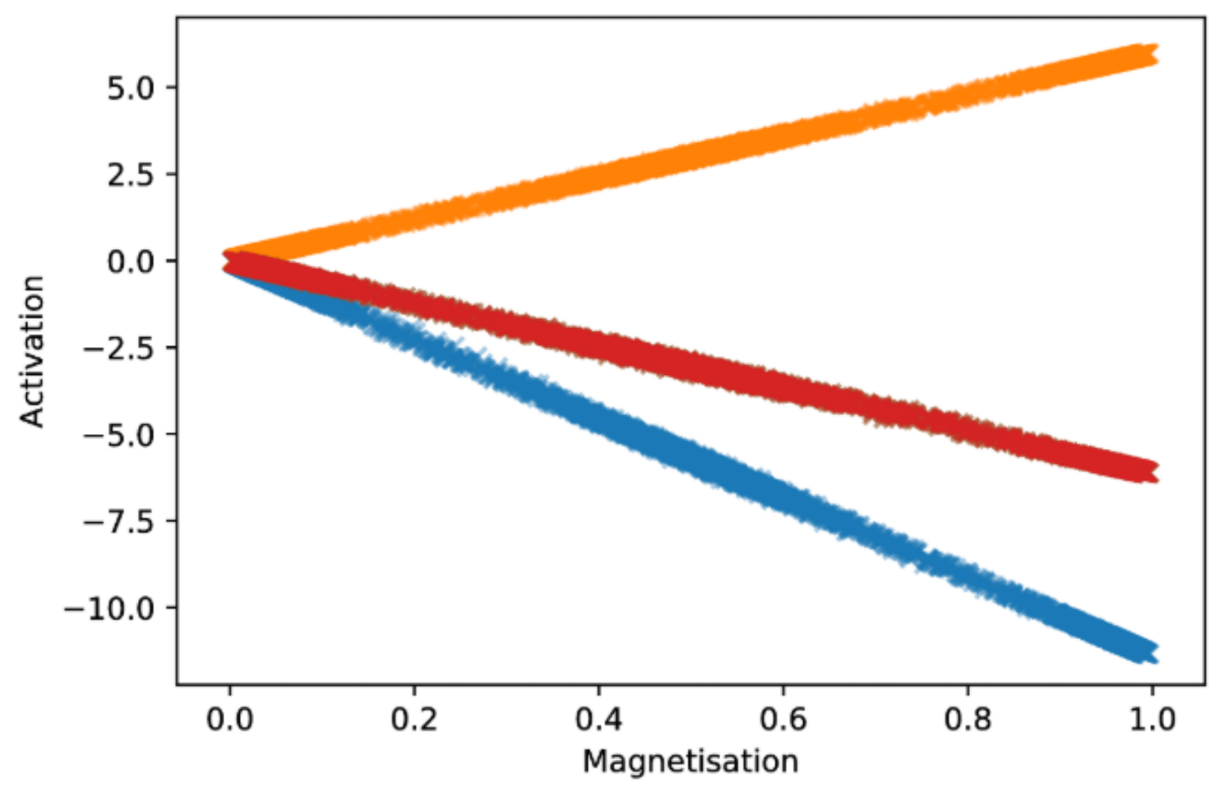


Figure 4: Activation function of the mean neurons associated with each cluster.

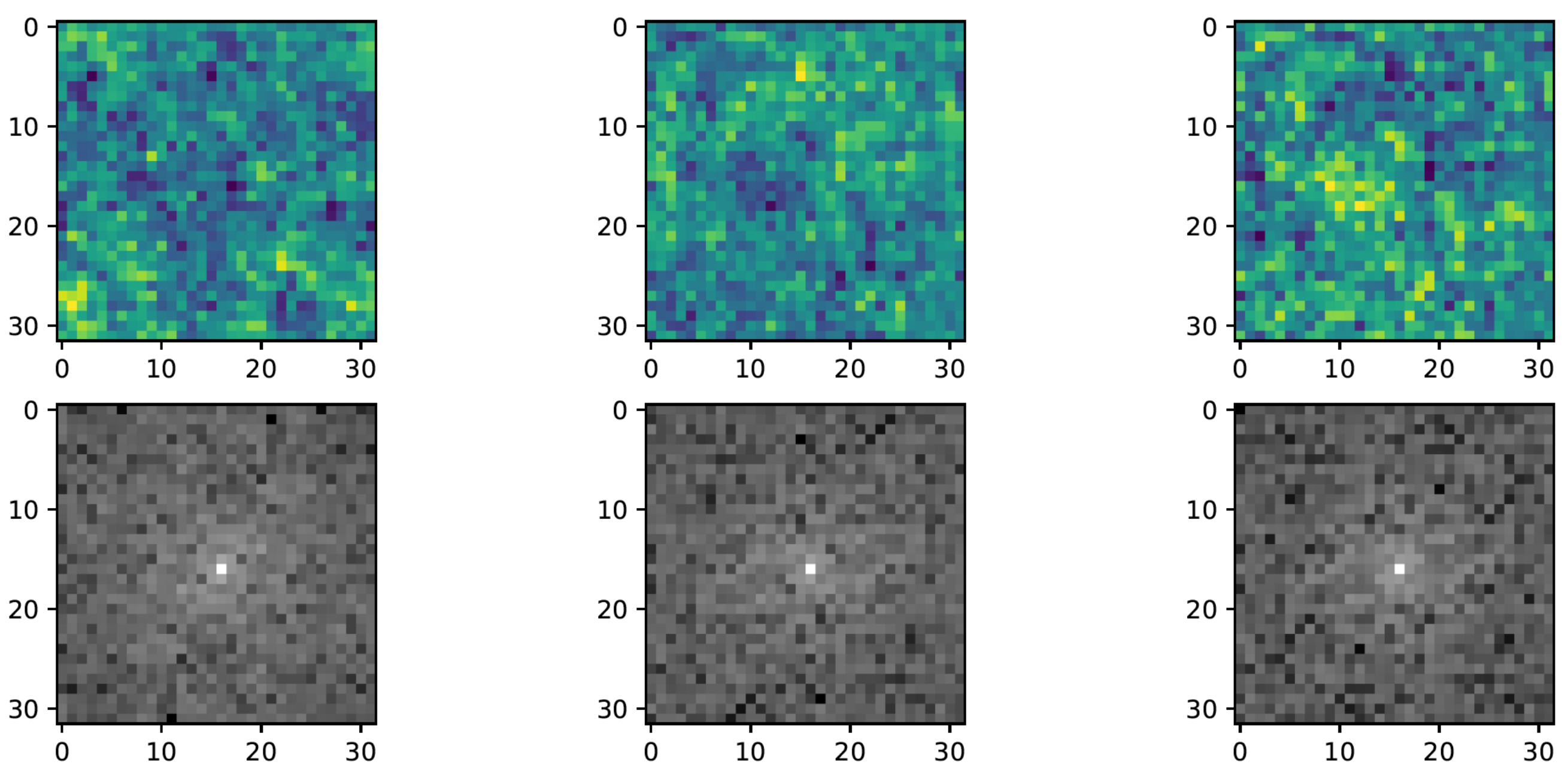


Figure 6: Weight matrices of the mean neurons associated with a three cluster classification.

4. Conclusions

- The toy models found on the literature [2, 4] emerge in large NN's as distinct classes of highly similar neurons.
- Two of the neuron classes learn to recognise the type of magnetisation. The third class, acts as an ambiguous unit.
- The featureless noise of the weight matrices is related to the symmetries of the Ising model [4]. The apparent negative-image inversion of two of the matrices suggests $\mathcal{Z}_2$ invariance
- The sharp increase of performance found at 3 hidden neurons might be explained by these classes.
- A certain amount of redundancy and randomness is beneficial to the performance of the network.

References

[1] Matthew J. S. Beach. Machine learning vortices at the Kosterlitz-Thouless transition. *Physical Review B*, 97(4), 2018.

[2] J Carrasquilla and R. G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, February 2017. arXiv: 1605.01735.

[3] D. Gomez. Physics monograph on Ising Model analysis based on KPCA: tarod13/Monograph, September 2018. original-date: 2018-08-16T16:14:42Z.

[4] P. Suchsland and S. Wessel. Parameter diagnostics of phases and phase transition learning by neural networks. *Physical Review B*, February 2018.

Escuela de Física-Matemática 2019