Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders

Sebastian J. Wetzel
Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
(Received 13 March 2017; revised manuscript received 9 June 2017; published 18 August 2017)

We examine unsupervised machine learning techniques to learn features that best describe configurations of the two-dimensional Ising model and the three-dimensional XY model. The methods range from principal component analysis over manifold and clustering methods to artificial neural-network-based variational autoencoders. They are applied to Monte Carlo–sampled configurations and have, *a priori*, no knowledge about the Hamiltonian or the order parameter. We find that the most promising algorithms are principal component analysis and variational autoencoders. Their predicted latent parameters correspond to the known order parameters. The latent representations of the models in question are clustered, which makes it possible to identify phases without prior knowledge of their existence. Furthermore, we find that the reconstruction loss function can be used as a universal identifier for phase transitions.

DOI: 10.1103/PhysRevE.96.022140

I. INTRODUCTION

Inferring macroscopic properties of physical systems from their microscopic description is an ongoing work in many disciplines of physics such as condensed matter, ultracold atoms, or quantum chromodynamics. The most drastic changes in the macroscopic properties of a physical system occur at phase transitions, which often involve a symmetry-breaking process. The theory of such phase transitions was formulated by Landau as a phenomenological model [1] and later devised from microscopic principles using the renormalization group [2,3]. Different phases can be identified by an order parameter that is zero in the disordered phase and nonzero in the ordered phase. Whereas in many known models the order parameter can be determined by symmetry considerations of the underlying Hamiltonian, there are states of matter where such a parameter can only be defined in a complicated nonlocal way [4]. These systems include topological insulators, quantum spin Hall states [5], or quantum spin liquids [6]. Therefore, it is important to develop new methods to identify parameters capable of describing phase transitions in these systems.

Such methods might be borrowed from machine learning. With the development of more powerful computers and artificial neural networks, machine learning has become one of the most influential disciplines of this century. It has been shown that such neural networks can approximate any continuous function under mild assumptions [7,8]. They quickly found applications in image classification [9], speech recognition [10], and natural language understanding [11].

In recent years physicists have started to employ machine learning techniques. Most of the tasks were tackled by supervised learning algorithms or with the help of reinforcement learning [12–24]. In supervised learning the algorithm is trained on labeled data to assign labels to data points. After successful training it can predict the labels of previously unseen data with high accuracy.

In addition to supervised learning, there exist unsupervised learning algorithms that can find structure in unlabeled data. It is already possible to employ unsupervised learning techniques to reproduce Monte Carlo–sampled states of the Ising model [25]. Phase transitions were found in an unsupervised manner using principal component analysis [26,27]. In this article we examine several other unsupervised learning algorithms such as manifold methods, clustering, and autoencoders. As a result of this examination we conclude that principal component analysis and variational autoencoders are the most promising among them to reveal phase transitions. This motivates us to transition to the employment of variational autoencoders and test how the latter handles different physical models. This algorithm finds a low-dimensional latent representation of the physical system that coincides with the correct order parameter. Furthermore, we find that autoencoders can reconstruct samples more accurately in the ordered phase, which suggests the use of the reconstruction error as a universal identifier for phase transitions.

Whereas for physicists this work is a promising way to find order parameters of systems where they are hard to identify, computer scientists and machine learning researchers might find an interpretation of the latent parameters.

II. MODELS

A. Ising model in two dimensions

The Ising model is one of the most studied and well understood models in physics. Whereas the one-dimensional classical Ising model does not possess a phase transition, the two-dimensional model does. The Hamiltonian of the Ising model on the square lattice with vanishing external magnetic h field reads

$$H(\mathbf{S}) = -J \sum_{\langle i,j\rangle_{\mathrm{NN}}} s_i s_j, \tag{1}$$

with uniform interaction strength J and spins $s_i \in \{+1 = \uparrow, -1 = \downarrow\}$ on each site $i = 1, \ldots, N$. The notation $\langle i,j\rangle_{\mathrm{NN}}$ indicates a summation over nearest neighbors. A spin configuration $\mathbf{S} = (s_1, \ldots, s_N)$ is a fixed assignment of a spin to each lattice site and Λ denotes the set of all possible configurations $\mathbf{S}$. We set the Boltzmann constant $k_B = 1$ and the interaction strength $J = 1$ for the ferromagnetic case and $J = -1$ for the antiferromagnetic case. A spin configuration $\mathbf{S}$ can be

expressed in matrix form as

$$\underline{\mathrm{S}} \hat{=} \begin{pmatrix} \uparrow & \downarrow & \uparrow & \cdots & \uparrow \\ \downarrow & \uparrow & \uparrow & \cdots & \uparrow \\ \vdots & \vdots & \vdots & & \vdots \\ \downarrow & \downarrow & \uparrow & \cdots & \downarrow \end{pmatrix}_{L\times L}. \tag{2}$$

Onsager solved the two-dimensional Ising model [28] and thereby calculated the critical temperature $T_c = 2/\ln(1+\sqrt{2}) = 2.269$.

For the purpose of this work, we assume a square lattice with length $L = 28$ such that $L \times L = N = 784$ and periodic boundary conditions. We sample the Ising model using a Monte Carlo algorithm [29] at temperatures $T \in [0,5]$ to generate 50 000 samples in the ferromagnetic case and 10 000 samples in the antiferromagnetic case. The Ising model obeys a discrete $\mathbb{Z}_2$ symmetry, which is spontaneously broken below T_c. The magnetization of a spin sample is defined as

$$M(\mathbf{S}) = \frac{1}{N}\sum_i s_i. \tag{3}$$

The partition function

$$Z = \sum_{\mathbf{S}\in\Lambda} \exp\left[-H(\mathbf{S})/T\right] \tag{4}$$

allows us to define the corresponding order parameter. It is the expectation value of the absolute value of the magnetization at fixed temperature

$$\langle \|M(T)\| \rangle = \frac{1}{Z}\sum_{\mathbf{S}\in\Lambda} \|M(\mathbf{S})\| \exp[-H(\mathbf{S})/T]. \tag{5}$$

Similarly, with the help of the matrix $A_{ij} = (-1)^{i+j}$, we define the order parameter of the antiferromagnetic Ising model, the expectation value of the staggered magnetization. The latter is calculated from an elementwise product with a matrix form of the spin configurations

$$M_{\mathrm{st}} = M(\underline{\mathrm{S}} \odot A). \tag{6}$$

B. The XY model in three dimensions

The Mermin-Wagner-Hohenberg theorem [30,31] prohibits continuous phase transitions in $d \leqslant 2$ dimensions at finite temperature when all interactions are sufficiently short ranged. Hence, we choose the XY model in three dimensions as a model to probe the ability of a variational autoencoder to classify phases of models with continuous symmetries. The Hamiltonian of the XY model reads

$$H(\mathbf{S}) = -J\sum_{\langle i,j\rangle_{\mathrm{NN}}} \mathbf{s}_i \cdot \mathbf{s}_j, \tag{7}$$

with spins on the one-sphere $\mathbf{s}_i \in \mathbb{R}^2$, $\|\mathbf{s}_i\| = 1$. Employing $J = 1$, the transition temperature of this model is $T_c = 2.2017$ [32]. Using a cubic lattice with $L = 14$ such that $N = L^3 = 2744$, we perform Monte Carlo simulations to create 10 000 independent sample spin configurations in the temperature range of $T \in [0,5]$. The order parameter is defined analogously to the Ising model magnetization (5), but with the L^2 norm of a magnetization consisting of two components.

III. METHODS

Principal component analysis [33] is an orthogonal linear transformation of the data to an ordered set of variables, sorted by their variance. The first variable, which has the largest variance, is called the first principal component, the variable with the second largest variance is the second principal component, and so on. The linear function $\langle\cdot,\mathbf{w}\rangle$, which maps a collection of spin samples $(\mathbf{S}_{(1)},\ldots,\mathbf{S}_{(n)})$ to its first principal component, is defined as

$$\arg\max_{\|\mathbf{w}\|=1}\left[\sum_j [(\mathbf{S}_{(j)} - \mu)\cdot\mathbf{w}]^2\right], \tag{8}$$

where μ is the vector of mean values of each spin averaged over the whole data set. Further principal components are obtained by subtracting the already calculated principal components and repeating Eq. (8) on the remaining subspace.

Kernel principal component analysis [34] projects the data into a kernel space in which the principal component analysis is then performed. In this work the nonlinearity is induced by a radial basis functions kernel.

Traditional neural-network-based autoencoders [35,36] consist of two artificial neural networks stacked on top of each other. They are created from an encoding artificial neural network, which outputs a latent representation of the input data, and a decoding neural network that tries to accurately reconstruct the input data from its latent representation (see Fig. 1). Very shallow versions of autoencoders can reproduce the results of principal component analysis [37]. The parameters of this algorithm are trained by performing gradient descent updates in order to minimize the reconstruction loss (reconstruction error) between input data and output data.

Variational autoencoders [38] are a modern version of autoencoders that impose additional constraints on the encoded representations, i.e., the latent variables in Fig. 1. These constraints transform the autoencoder to an algorithm that learns a latent variable model for its input data. Whereas the neural networks of traditional autoencoders learn an arbitrary function to encode and decode the input data, variational autoencoders learn the parameters of a probability distribution modeling the data. After learning the probability distribution, one can sample parameters from it and then let the encoder network generate samples closely resembling the training data. To achieve this, variational autoencoders employ the assumption that one can sample the input data from a unit Gaussian distribution of latent parameters. The weights of the model are trained by simultaneously optimizing two loss functions, a reconstruction loss and the Kullback-Leibler

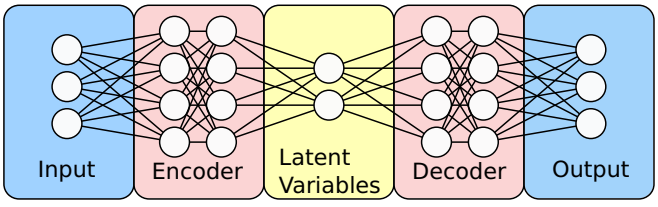

FIG. 1. Autoencoder neural network architecture. The encoder network translates the input to its latent representation, from which the decoder reconstructs an approximation of the input as output.

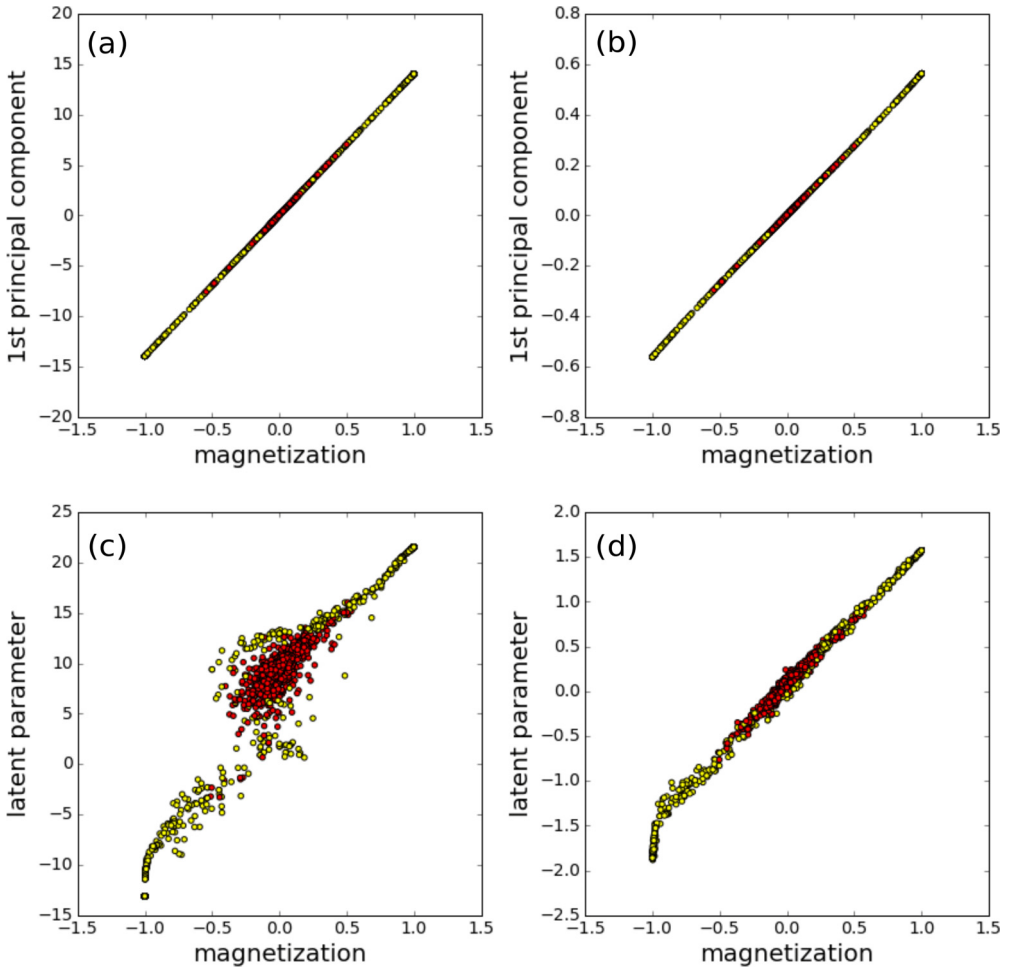

FIG. 2. Ferromagnetic Ising model showing the principal components and latent representations versus magnetization for different algorithms: (a) principal component analysis, (b) kernel principal component analysis, (c) autoencoder, and (d) variational autoencoder. Red points correspond to configurations of the unordered phase, while yellow points belong to the ordered phase.

divergence between the learned latent distribution and a prior unit Gaussian.

In this work we use autoencoders and variational autoencoders with one fully connected hidden layer in the encoder as well as one fully connected hidden layer in the decoder, each consisting of 256 neurons. The number of latent variables is chosen to match the model from which we sample the input data. The activation functions of the intermediate layers are rectified linear units. The activation function of the final layer is a sigmoid in order to predict probabilities of spin $\uparrow$ or $\downarrow$ in the Ising model or tanh for predicting continuous values of spin components in the XY model. We do not employ any L^1, L^2, or dropout regularization. However, we tune the relative weight of the two loss functions of the variational autoencoder to fit the problem at hand. The Kullback-Leibler divergence of the variational autoencoder can be regarded as regularization of the traditional autoencoder. In our autoencoder the reconstruction loss is the cross-entropy loss between the input and output probability of discrete spins, as in the Ising model. The reconstruction loss is the mean-square error between the input and the output data of continuous spin variables in the XY model.

To understand why a variational autoencoder can be a suitable choice for the task of classifying phases, we recall what happens during training. The weights of the autoencoder learn two things. On the one hand, they learn to encode the similarities of all samples to allow for an efficient reconstruction. On the other hand, they learn a latent distribution of the parameters that encode the most information possible to distinguish between different input samples. Let us translate these considerations to the physics of phase transitions. If all the training samples are in the unordered phase, the autoencoder learns the common structure of all samples. The autoencoder fails to learn any random entropy fluctuations, which are averaged out over all data points. However, in the ordered phase there exists a common order in samples belonging to the same phase. This common order translates to a nonzero latent parameter, which encodes correlations on each input sample. It turns out that in our cases this parameter is the order parameter corresponding to the broken symmetry. It is not necessary to find a perfect linear transformation between the order parameter and the latent parameter as is the case in Fig. 2. A one-to-one correspondence is sufficient, such that one is able to define a function that maps these parameters onto each other and captures all discontinuities of the derivatives of the order parameter.

Principal component analysis and autoencoders seem very different, but they share common characteristics.

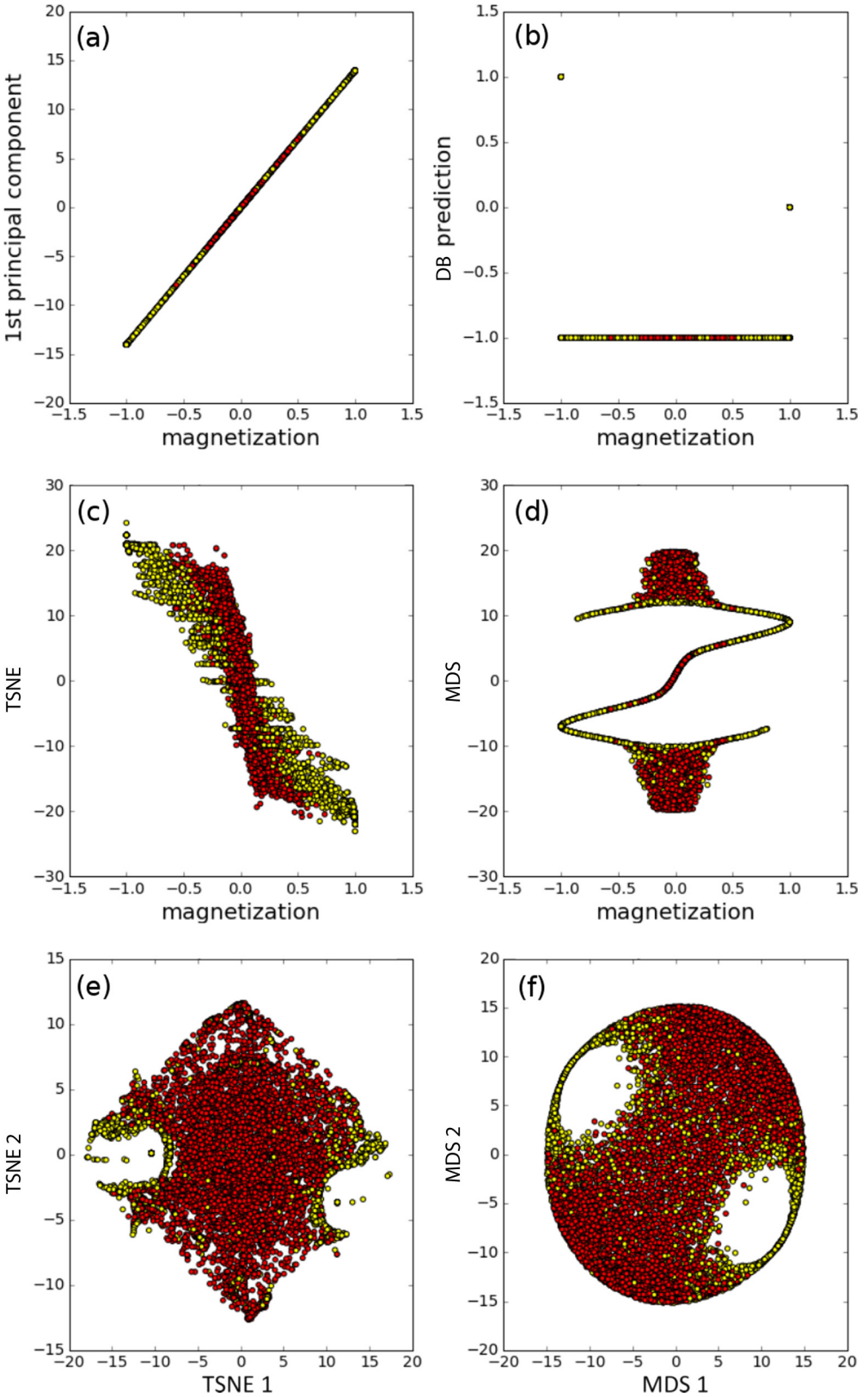

FIG. 3. Ferromagnetic Ising model showing a comparison of manifold and clustering methods: (a) PCA, (b) DBSCAN, (c) one-component TSNE, (d) one-component MDS, (e) two-component TSNE, and (f) two-component MDS.

Reconstructing the input data from its principal components minimizes the mean-square reconstruction error. Hence, a single-layer autoencoder with linear activation functions closely resembles principal component analysis [37]. Principal component analysis is much easier to apply and in general is characterized by fewer parameters than autoencoders. Autoencoders based on convolutional layers can have a reduced number of parameters. In extreme cases this number can be even less than the parameters of principal component analysis. Furthermore, such autoencoders can promote locality of features in the data.

We further examine manifold methods, where high-dimensional data are embedded in a low-dimensional manifold in which similar data points are represented close to each other. For this purpose we employ t-distributed stochastic neighbor embedding (TSNE) [39], a technique that is particularly sensitive to local structure. We also consider multidimensional scaling (MDS) [40], which seeks a low-dimensional representation of the data in which the distances respect the distances in the original high-dimensional space. In the following sections we embed the spin configurations into a one- (one-component TSNE or MDS) or two-dimensional manifold (two-component TSNE or MDS).

As an example of clustering methods we try to employ density-based spatial clustering of applications with noise (DBSCAN) [41]. It associates clusters with areas of high density separated by areas of low density. The DBSCAN can also find convex-shaped clusters.

For this work we employed the PYTHON libraries sklearn [42] and keras [43]. A detailed introduction to autoencoders can be found in [44].

IV. RESULTS

A. Ferromagnetic Ising model

First we compare the four most successful algorithms applied to the Ising model: principal component analysis (PCA), kernel principal component analysis, autoencoders, and variational autoencoders. They all share the characteristic that the first principal components or the latent parameters (Fig. 2) show a clear correlation to the magnetization. However, the traditional autoencoder fails to capture this correlation in the vicinity of magnetization $M(S) = 0$; this fact lets us favor variational autoencoders over traditional autoencoders. The principal component methods show the most accurate results, slightly better than the variational autoencoder. This is to be expected, since the former are modeled by fewer parameters.

In Fig. 3 we compare manifold and clustering methods to principal component analysis. All algorithms are employed on raw data in the temperature range $T \in [0,5]$. The PCA is the only method that successfully manages to approximate the magnetization. The DBSCAN completely fails to find clusters on raw data. Moreover, the one-component TSNE can separate clusters of positive and negative magnetization. One-component MDS finds a structure that separates the ordered phase (yellow) from the unordered phase (red). However, this structure cannot be interpreted in a physical sense. Two-component TSNE and two-component MDS both successfully place samples belonging to the same phases close to each other. They also distinguish between two sorts of magnetization. We expect that in the case of supervised learning, where one can cluster the training data, one would obtain much clearer results.

In the following results we concentrate on the variational autoencoder as the most promising and powerful algorithm for unsupervised learning.

As a starting point, we choose the number of latent parameters in the variational autoencoder to be one. After training the network for 50 epochs and observing a saturation of the training loss, we visualize the results in Fig. 4. In Fig. 4(a) we see a close linear correlation between the latent parameter and the magnetization. In Fig. 4(b) there is a histogram of spin configurations encoded into their latent parameter. The model learned to classify the configurations into three clusters. The identification of the latent parameter as a close approximation of the magnetization $M(S)$ allows

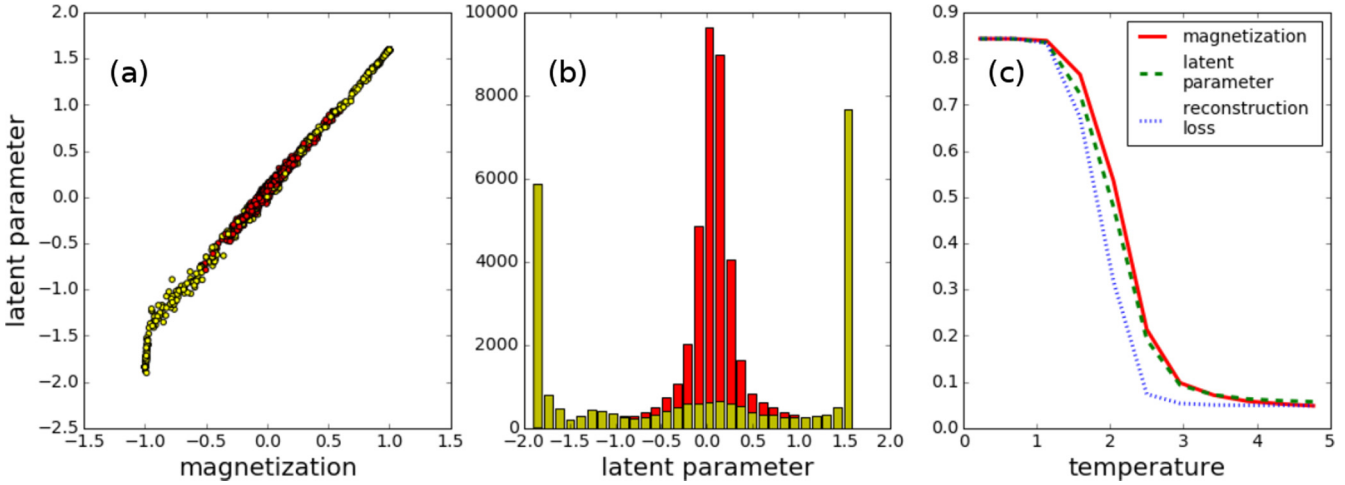

FIG. 4. Ferromagnetic Ising model. (a) The correlation between the latent parameter and magnetization is shown for each spin sample. Red dots indicate points in the unordered phase, while yellow dots correspond to the ordered phase. (b) The histogram counts occurrences of latent parameters. (c) One can see the average values at fixed temperature of the absolute value of magnetization, the absolute value of the latent parameter, and the cross-entropy reconstruction loss. The reconstruction loss is mapped on the $T = 0$ and 5 values of the magnetization and the latent parameter is rescaled to the magnetization at $T = 0$.

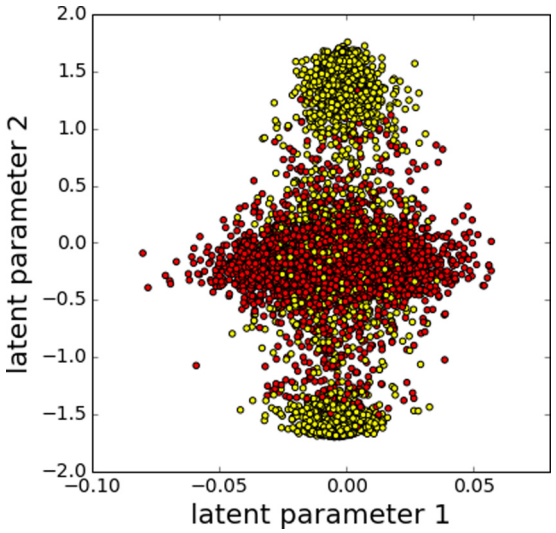

FIG. 5. Ferromagnetic Ising model with visualization of data in a two-dimensional latent space. Red dots indicate points in the unordered phase, while yellow dots correspond to the ordered phase. The axis for parameter 1 has a smaller range than the axis for parameter 2.

us to interpret the properties of the clusters. The right and left clusters in Fig. 4(b) correspond to an average magnetization of $M(S) \approx \pm 1$, while the middle cluster corresponds to the magnetization $M(S) \approx 0$. Employing a different viewpoint, from Fig. 4 we conclude that the parameter that holds the most information on how to distinguish Ising spin samples is the order parameter. In Fig. 4(c) the average of the magnetization, the latent parameter, and the reconstruction loss are shown as a function of the temperature. A sudden change in the magnetization at $T_c \approx 2.269$ defines the phase transition between paramagnetism and ferromagnetism. Even without previous knowledge of this order parameter, we can use the results of the autoencoder to infer the position of the phase transition. As an approximate order parameter, the average absolute value of latent parameter also shows a steep change at T_c. The averaged reconstruction loss also changes drastically at T_c. While the latent parameter is different for each physical model, the reconstruction loss can be used as a universal parameter to identify phase transitions. In conclusion, without any knowledge of the Ising model and its order parameter but sample configurations, we can find a good estimation for its order parameter and the occurrence of a phase transition.

It is *a priori* not clear how to determine the number of latent neurons in the creation of the neural network of the autoencoder. Due to the lack of theoretical groundwork, we find the optimal number by experimentation. If we expand the number of latent dimensions by one (see Fig. 5), the results of our analysis change only slightly. The second parameter contains much less information compared to the first, since it stays very close to zero. Hence, for the Ising model, one parameter is sufficient to store most of the information of the latent representation.

B. Antiferromagnetic Ising model

After having identified variational autoencoders as the most promising unsupervised learning algorithms to determined phase transitions, we concentrate on this algorithm in the following and present the results of the other algorithms in the Appendix.

While the ferromagnetic Ising model serves as an ideal starting ground, in the next step we are interested in models where different sites in the samples contribute in a different manner to the order parameter. We do this in order to show that our model is even sensitive to structure on the smallest scales. For the magnetization in the ferromagnetic Ising model, all spins contribute with the same weight. In contrast, in the antiferromagnetic Ising model, neighboring spins contribute with opposite weight to the order parameter (6).

Again the variational autoencoder manages to capture the traditional order parameter. The staggered magnetization is strongly correlated with the latent parameter (see Fig. 6). The three clusters in the latent representation make it possible to

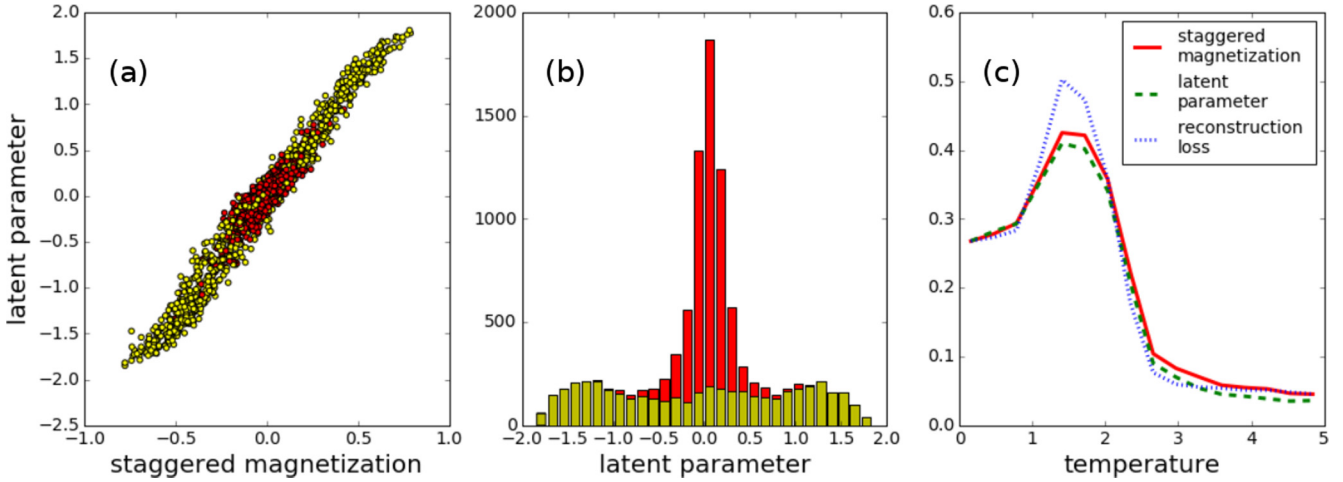

FIG. 6. Antiferromagnetic Ising model. (a) Correlation between the latent parameter and staggered magnetization for each spin sample. Red dots indicate points in the unordered phase, while yellow dots indicate points in the ordered phase. (b) The histogram counts occurrences of latent parameters. (c) Average at fixed temperature of the absolute value of staggered magnetization, the absolute value of the latent parameter, and the cross-entropy reconstruction loss. The reconstruction loss is mapped on the $T = 0$ and 5 values of the staggered magnetization and the latent parameter is rescaled to the magnetization at $T = 0$.

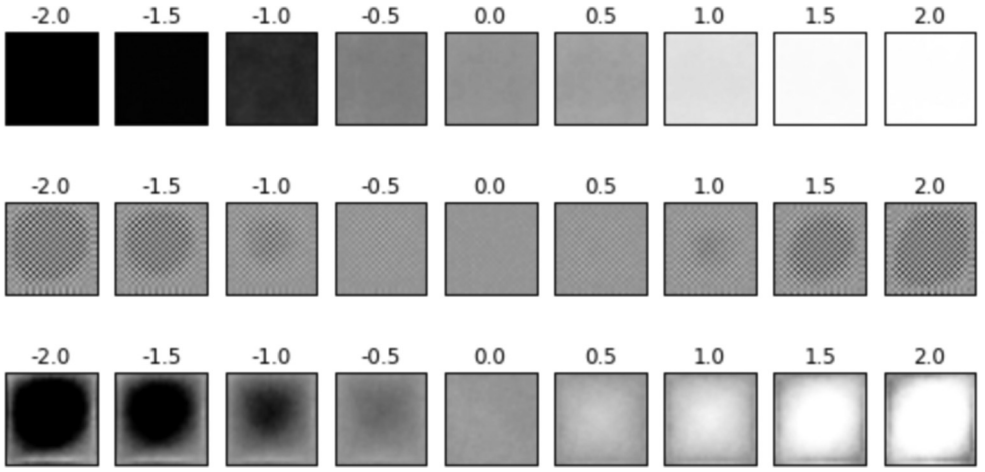

FIG. 7. Reconstruction of images, each consisting of 28×28 pixels, from the latent parameter. The brightness indicates the probability of the spin to be up [white indicates $p(\uparrow) = 1$ and black $p(\downarrow) = 1$]. The first row is a reconstruction of sample configurations from the ferromagnetic Ising model. The second row corresponds to the antiferromagnetic Ising model. The third row is the prediction from the AF latent parameter, where each second spin is multiplied by -1, to show that the second row indeed predicts an antiferromagnetic state.

interpret different phases. Furthermore, we note that all three averaged quantities, the magnetization, the latent parameter, and the reconstruction loss, can serve as indicators of a phase transition.

Figure 7 demonstrates the reconstruction from the latent parameter. In the first row we see the reconstruction from samples of the ferromagnetic Ising model; the latent parameter encodes the whole spin order in the ordered phase. Reconstructions from the antiferromagnetic Ising model are shown in the second and third rows. Since the reconstructions clearly show an antiferromagnetic phase, we infer that the autoencoder encodes the spin samples even to the most microscopic level.

C. The XY model

In the XY model we examine the capabilities of a variational autoencoder to encode models with continuous symmetries. The application of other algorithms is compared in the Appendix. In models like the Ising model, where discrete symmetries are present, the autoencoder only needs to learn a discrete set, which is often finite, of possible representations of the symmetry-broken phase. If a continuous symmetry is broken, there are infinitely many possibilities of how the ordered phase can be realized. Hence, in this section we test the ability of the autoencoder to embed all these different realizations into latent variables.

The variational autoencoder handles this model equally well as the Ising model. We find that two latent parameters model the phase transition best. The latent representation in Fig. 8(b) shows the distribution of various states around a central cluster. The radial symmetry in this distribution leads to the assumption that a sensible order parameter is constructed from the L^2 norm of the latent parameter vector. In Fig. 8 one sees the correlation between the magnetization and the

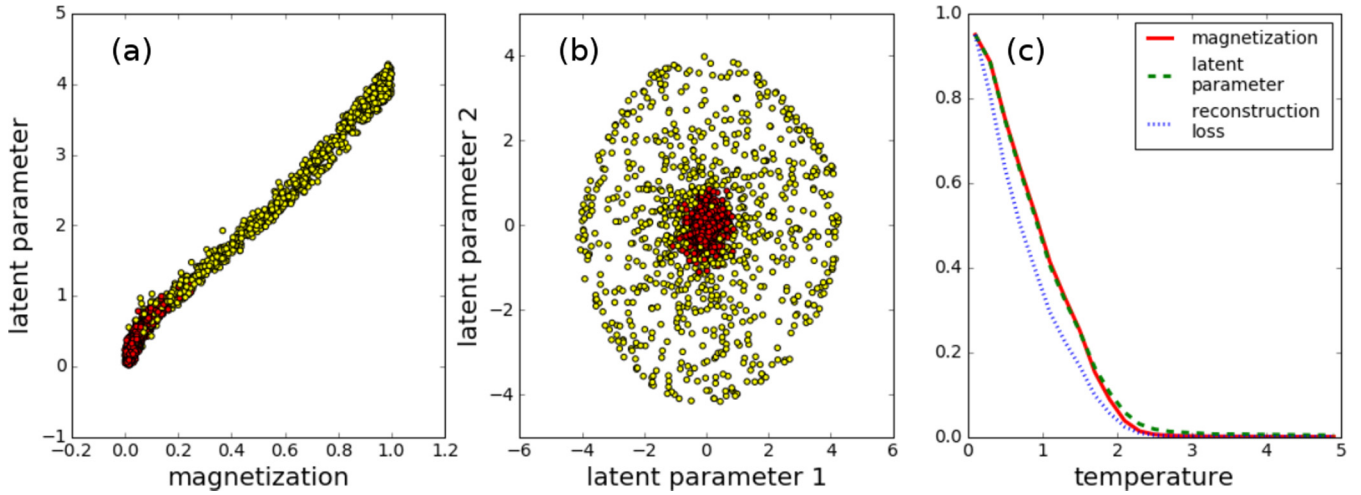

FIG. 8. The XY model. (a) Correlation between the L^2 norm of the latent parameter vector and the L^2 norm of the magnetization for each spin sample. Red dots indicate points in the unordered phase, while yellow dots indicate points in the ordered phase. (b) Representation of the spin configurations in two-dimensional latent space. (c) For each L^2 norm of the magnetization, the L^2 norm of latent parameter, and the average of the square root of the mean-square error reconstruction loss, we plot the average at fixed temperature. The reconstruction loss is mapped on the $T = 0$ and 5 values of the magnetization and the latent parameter is rescaled to the magnetization at $T = 0$.

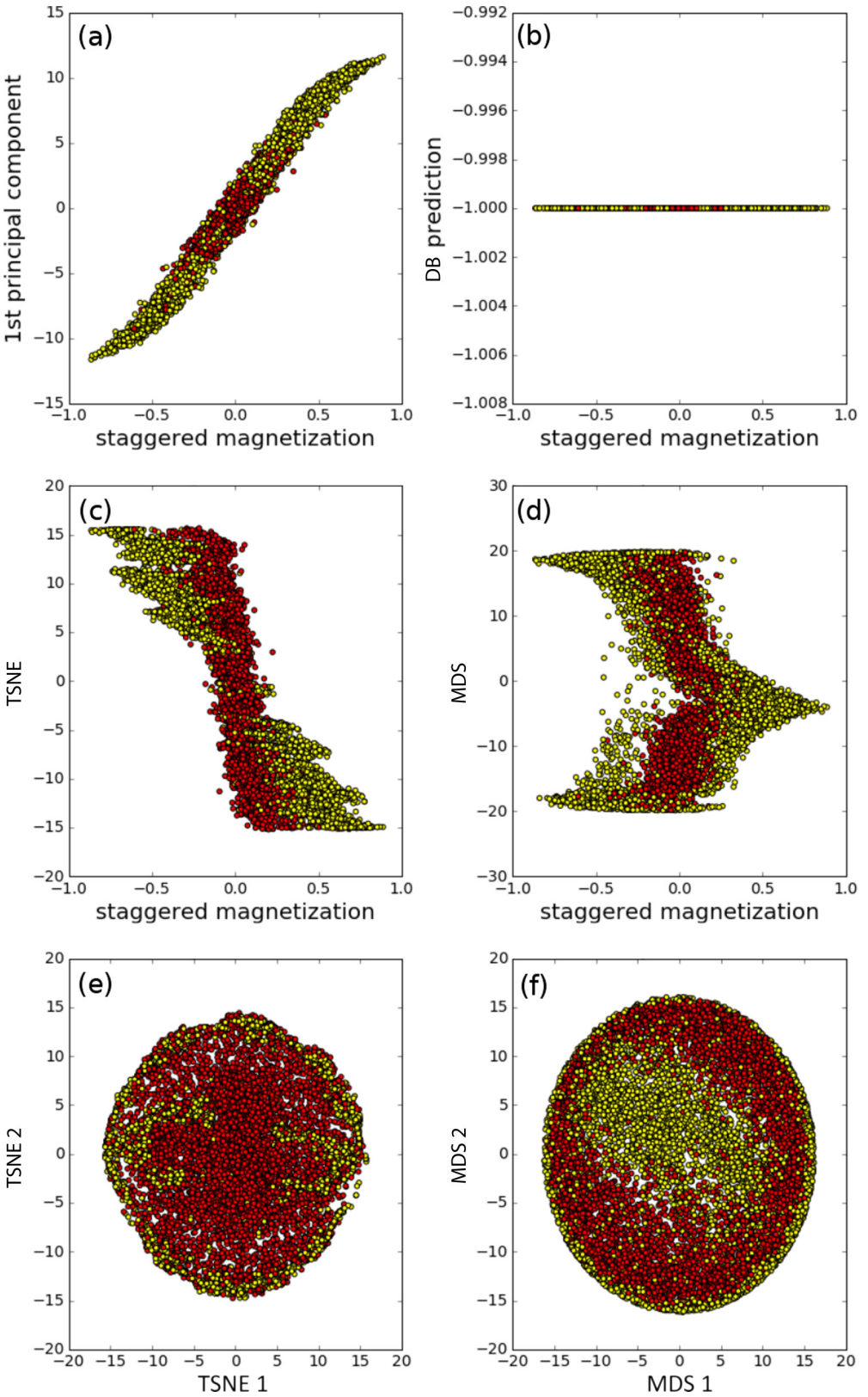

FIG. 9. Antiferromagnetic Ising model showing a comparison of the manifold and clustering methods: (a) PCA, (b) DBSCAN, (c) one-component TSNE, (d) one-component MDS, (e) two-component TSNE, and (f) two-component MDS.

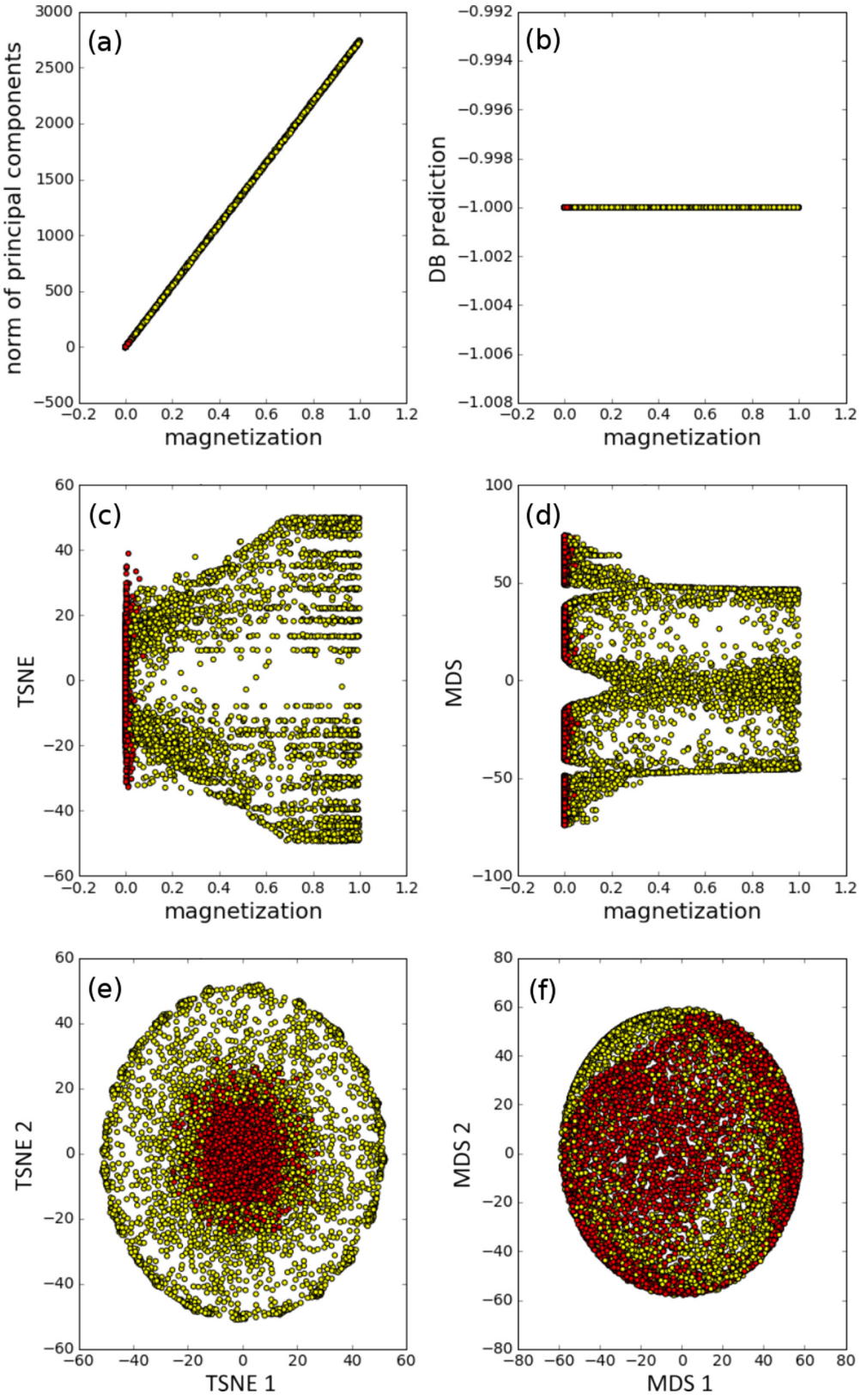

FIG. 10. The XY model showing a comparison of the manifold and clustering methods: (a) PCA, (b) DBSCAN, (c) one-component TSNE, (d) one-component MDS, (e) two-component TSNE, and (f) two-component MDS.

absolute value of the latent parameter vector. Averaging the samples for the same temperature hints at the fact that the latent parameter and the reconstruction loss can serve as an indicator for the phase transition.

V. CONCLUSION

We have shown that it is possible to observe phase transitions using unsupervised learning. We compared different unsupervised learning algorithms and found that principal component analysis and variational autoencoders are the best algorithms for examining phase transitions. We were motivated by the need for an upgrade of the traditional autoencoder to a variational autoencoder. The weights and latent parameters of the variational autoencoder are able to store information about microscopic and macroscopic properties of the underlying systems. The most distinguished latent parameters coincide with the known order parameters. Furthermore, we have established the reconstruction loss as a universal indicator for phase transitions. We expanded the toolbox of unsupervised learning algorithms in physics by powerful methods, most notably the variational autoencoder, which can handle nonlinear features in the data and scale very well to huge data sets. In the future one may employ autoencoders to capture phase transitions with nonlinear order parameters, which PCA cannot reproduce. These theories include lattice gauge theories, where the order parameters are defined by a loop along several lattice sites. We expect the prediction of unseen phases or exposure of unknown order parameters, e.g., in quantum spin liquids. We look forward to the development of deep convolutional autoencoders that have a reduced number of parameters compared to fully connected autoencoders to probe locality in feature selection. Furthermore, since there exists a connection between deep neural networks and renormalization group [45], it may be helpful to employ deep convolutional autoencoders to further expose this connection.

ACKNOWLEDGMENTS

The author would like to thank Timo Milbich, Björn Ommer, Michael Scherer, Manuel Scherzer, and Christof Wetterich for useful discussions as well as Mathias Neidig and Shirin Nkongolo for helpful suggestions. S.W. acknowledges support from the Heidelberg Graduate School of Fundamental Physics.

APPENDIX: PCA, MANIFOLD, AND CLUSTERING RESULTS

1. Antiferromagnetic Ising model

In Fig. 9 we compare different unsupervised learning algorithms applied to the antiferromagnetic Ising model. Principal component analysis approximates the magnetization by its first principal component equally well as variational autoencoders (Fig. 6). The DBSCAN fails to find any meaningful structure. One-component TSNE finds two clusters of staggered magnetizations. One-component MDS is able to discover structure in the data, however we cannot relate it to any physical quantity. Two-component TSNE projects the ordered phase (yellow) to the outside of the manifold, while the unordered phase (red) is mapped to the inside. It also finds substructure in the unordered phase. Two-component MDS is able to separate points belonging to the unordered phase from points belonging to the ordered phase.

2. The XY model

Figure 10 compares the results of different unsupervised learning algorithms applied to raw Monte Carlo simulations of the XY model. The Euclidean norm of the first two principal components analysis is perfectly correlated with the magnetization. Principal component analysis thus yields similar results as variational autoencoders (Fig. 8). One-component TSNE (one-component MDS) finds a structure where two (three) points in the embedded manifold correspond to a finite magnetization. The most promising result is achieved by two-component TSNE: Spin configurations belonging to the unordered phase are centered at the origin, while configurations belonging to the ordered phase are mapped to the exterior. This result is similar to the clustering from the variational autoencoder (see Fig. 8). Two-component MDS finds two clusters of a higher density of data points belonging to the ordered phase; the result is less pronounced compared to TSNE.

[1] L. Landau, Phys. Z. Sowjetunion **11**, 26 (1937).
[2] L. P. Kadanoff, Physics **2**, 263 (1966).
[3] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
[4] X.-G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, Oxford, 2004).
[5] C.-C. Kaun and T. Seideman, Phys. Rev. Lett. **94**, 226801 (2005).
[6] P. Anderson, Mater. Res. Bull. **8**, 153 (1973).
[7] G. Cybenko, Math. Control Signals Syst. **2**, 303 (1989).
[8] K. Hornik, Neural Networks **4**, 251 (1991).
[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran, Red Hook, 2012), pp. 1097–1105.
[10] G. Hinton *et al.*, IEEE Signal Process. Mag. **29**, 82 (2012).
[11] Y. Goldberg, arXiv:1510.00726.
[12] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, Phys. Rev. Lett. **91**, 135503 (2003).
[13] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[14] Z. Li, J. R. Kermode, and A. D. Vita, Phys. Rev. Lett. **114**, 096405 (2015).
[15] E. LeDell, Prabhat, D. Y. Zubarev, B. Austin, and W. A. Lester, Jr., J. Math. Chem. **50**, 2043 (2012).
[16] G. Pilania, J. E. Gubernatis, and T. Lookman, Phys. Rev. B **91**, 214302 (2015).
[17] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky, and W. Andreoni, Phys. Rev. B **85**, 104104 (2012).

[18] O. S. Ovchinnikov, S. Jesse, P. Bintacchit, S. Trolier-McKinstry, and S. V. Kalinin, Phys. Rev. Lett. **103**, 157203 (2009).
[19] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[20] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Phys. Rev. Lett. **108**, 253002 (2012).
[21] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater. **22**, 3762 (2010).
[22] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).
[23] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, JHEP 05 (2017) 006.
[24] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[25] G. Torlai and R. G. Melko, Phys. Rev. B **94**, 165134 (2016).
[26] L. Wang, Phys. Rev. B **94**, 195105 (2016).
[27] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).
[28] L. Onsager, Phys. Rev. **65**, 117 (1944).
[29] N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44**, 335 (1949).
[30] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).
[31] P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).
[32] A. P. Gottlob and M. Hasenbusch, Physica A **201**, 593 (1993).
[33] K. Pearson, Philos. Mag. Series 6 **2**, 559 (1901).
[34] B. Scholkopf, A. Smola, and K.-R. Müller, *Advances in Kernel Methods* (MIT Press, Cambridge, 1999), pp. 327–352.
[35] H. Bourlard and Y. Kamp, Biol. Cybern. **59**, 291 (1988).
[36] G. E. Hinton and R. S. Zemel, in Advances in *Neural Information Processing 6*, edited by J. D. Cowan, G. Tesauro, and J. Alspector (Kaufmann, San Mateo, 1994).
[37] P. Baldi and K. Hornik, Neural Networks **2**, 53 (1989).
[38] D. P. Kingma and M. Welling, arXiv:1312.6114.
[39] L. van der Maaten and G. Hinton, J. Mach. Learn. Res. **9**, 2579 (2008).
[40] J. Kruskal, Psychometrika **29**, 115 (1964).
[41] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, edited by E. Simoudis, J. Han, and U. Fayyad (AAAI, Palo Alto, 1996), pp. 226–231.
[42] F. Pedregosa *et al.*, J. Mach. Learn. Res. **12**, 2825 (2011).
[43] F. Chollet, keras, 2015, https://github.com/fchollet/keras.
[44] F. Chollet, Building autoencoders in keras, 2014, https://blog.keras.io/building-autoencoders-in-keras.html.
[45] P. Mehta and D. J. Schwab, arXiv:1410.3831.