DISCOVERING PHASES AND PHASE TRANSITIONS USING MACHINE LEARNING

by

Jairo Jose Orozco Sandoval

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

PHYSICS

UNIVERSITY OF PUERTO RICO

MAYAGÜEZ CAMPUS

2019

Approved by:

Rafael Ramos, Ph.D. President, Graduate Committee	Date
Hector Jimenez, Ph.D. Member, Graduate Committee	Date
Kejie Lu, Ph.D. Member, Graduate Committee	Date
Yang Li, Ph.D. Member, Graduate Committee	Date
Hilton Alers, Ph.D. Representative of Graduate Studies	Date
Rafael Ramos, Ph.D. Chairperson of the Department	Date

ABSTRACT

Machine learning a specific subset of artificial intelligence, trains a machine to learn from data. It has become a robust method for the identification of patterns within complex physical systems to determine certain physical quantities without prior knowledge of their physics principles. In this thesis work, we apply an unsupervised machine learning technique, Principal Component Analysis, and a supervised learning technique, Artificial Neural Networks, to identify phases and phase transitions in square and hexagonal lattice Ising models. It can be drawn from the results that Principal Component Analysis can successfully identify phase transitions and locate the transition temperatures in both square and hexagonal lattice systems. Additionally, it was found that two principal components are related to the order parameter and the susceptibility of the systems. The weight vectors have, then, a physical explanation, which is helpful to better understand system behavior. On the other hand, by employing neural networks, it was possible to understand the training effects of the Ising model, as well as obtain a critical temperature value T_c close to the real thermodynamic value. This confirms machine learning is suitable for approaching the type of complex systems studied in this research. Principal Component Analysis and neural network can learn from complex data, without the need of significant human intervention.

RESUMEN

El aprendizaje automático, un subconjunto específico de inteligencia artificial, se entrena a la máquina para aprender de datos no explorados. Se ha convertido en un método robusto para la identificación de patrones dentro de sistemas físicos complejos para determinar ciertas cantidades físicas sin el conocimiento previo de sus principios físicos. En este trabajo de tesis, aplicamos una técnica de aprendizaje automático no supervisado, el Análisis de Componentes Principales (ACP) y una técnica de aprendizaje supervisado, Redes Neuronales Artificiales (RNA), para identificar fases y transiciones de fase en modelos de Ising de red cuadrada y hexagonal. Se puede deducir de los resultados que el análisis de componentes principales puede identificar con éxito las transiciones de fase y ubicar las temperaturas de transición tanto en los sistemas de red cuadrada como en los hexagonales. Además, se encontró que los primeros componentes principales están relacionados con el parámetro de orden y la susceptibilidad de los sistemas. Los vectores de peso tienen, entonces, una explicación física, que es útil para comprender mejor el comportamiento del sistema. Por otro lado, al emplear redes neuronales, fue posible comprender los efectos de entrenamiento del modelo de Ising, así como obtener un valor crítico de temperatura T_c cercano al valor termodinámico real. Esto confirma que el aspecto iterativo del aprendizaje automático lo hace adecuado para abordar el tipo de sistemas complejos estudiados en esta investigación. Si los modelos ACP y RN están expuestos a datos complejos, se adaptan y aprenden de iteraciones anteriores, produciendo resultados que son confiables y repetibles, sin la necesidad de una intervención humana significativa.

Dedicado a Dios y a mi esposa.

ACKNOWLEDGEMENTS

- First, I want to thank God for allowing me to fulfill another dream in my life, and that without Him this would not have been possible.

- Thank to my wife, Karla Echeverria for her love, support, understanding and help in all this time, because she was always with me and this work is dedicated to her.

- To my parents jairo and Esperanza and my siblings Jose and Lisbeth for being a support in our lives and in all this time that we have been here in Puerto Rico.

- To Dr. Lu and Dr. Ramos for their help and cooperation for the realization of this work.

- Thanks to Alejandro Rigau for his great help in the calculations of Neural Network and for his constant interest in always giving me a hand for this work.

- To my friends Camilo, Kevin, Larry, Mileydis, Marina, Wilito, Sonalí , Joseph, Joselito and Valerie for their companionship, friendship and emotional support that helped me to carry out this work.

- To the University of Puerto Rico for opening the doors and letting me be part of this great institution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

In recent years, machine learning has been a helpful tool to discover new physics without prior human knowledge, specifically in complex problems in physics. This means we can know how a physical system behaves even without knowing its Hamiltonian. It has been possible to use machine learning to predict crystal structures [1.2], approximate density functions [1.3], model molecular atomization energy [1.4], and many other applications. Machine learning enables computers to learn from various experiences, events or data; once the computer learns, generalizes the knowledge learned to later solve problems without knowing the basis of the problem. Machine learning has had a great impact, through the different supervised and unsupervised machine learning techniques. In supervised machine learning, we have a data set and we already know what our correct outputs should look like. There is a relationship between the input and the output, this output is called label. Unsupervised machine learning allows us to approach problems with no idea of what our results should look like; we can make predictions of a certain problem without knowing the labels by clustering the data, based in the relationship among the variables in the data.

For many years classifying and discovering phases and phase transitions is one of the most important topics in Condensed Matter Physics [1.1]; however, it is not an easy job to do, especially when we work with complex systems and the number of states is very large. One of the most important models in the theoretical physics to study phase transitions is the Ising model, given it was the first model that could successfully predict a phase transition [1.5,1.6]. The Ising model models a magnet; it is represented in a lattice and each site of the

lattice is occupied by a "spin," an arrow pointing up or down. These spins model the unpaired electrons in the atoms that have magnetic moments, for example the iron atoms. The lattice models the fact, that the atoms are in a crystal, with a regular structure. The Ising model has been studied in square and triangular lattices, but very rarely in hexagonal lattice. One of the most common methods used to study the Ising model is the Monte Carlo method. This method is used to calculate numerically the thermodynamic properties as it averages in a system by numerical simulations. The idea of the method is to find an algorithm to generate a long sequence of configurations of a system, such that after a while, each configuration is generated with the appropriate probability to describe the equilibrium of the system [1.8]. Recently, the Monte Carlo method has been used together with the machine learning techniques [1.1, 1.8, 1.7] for the study and discovery of phases and phase transitions. The Monte Carlo method is used for the creation of data to later implement the machine learning techniques.

Motivated by the search of physical properties through the machine learning techniques, in this Thesis work, we apply unsupervised machine learning technique, the Principal Component Analysis (PCA), and supervised learning technique, the feed-forward Neural Network (NN), to study square and hexagonal ferromagnetic lattice Ising systems, in order to recognize phases, phase transitions, and related physical properties without knowing any information about the microscopic theory or the order parameters. We study the Ising model in square lattice as our toy model, once it was studied, we apply PCA in hexagonal lattice and later compare the results. The interest in the hexagonal ferromagnetic system is because this system has been rarely studied. The literature has only studied the square

ferromagnetic and triangular antiferromagnetic systems applying machine learning techniques.

We also create a Feed-forward NN based on [4.8] and guided by Carrasquilla et al. [4.2], to show the training effects of the neural network in the Ising model using only two neurons. Also, a NN with 300 neurons in the hidden layer was employed to get a good accuracy of phase prediction in hexagonal lattice.

By means of these machine learning techniques, in this work, PCA was able to recognize phase transitions in both systems, to make a good approximation of the critical temperature T_c, and to observe the behavior of the magnetization (order parameter) and susceptibility. With neural networks, the critical temperature and the order parameter were also identified in the hexagonal system, making it possible to understand the training effect of the Ising model in the neural network and observe that the magnetization of the system is encoded in the hidden layer. In this thesis work we will prove that PCA and NN can successfully find physics properties in the Ising model, corroborating the efficiency of these machine learning methods.

This document is organized as follows: Chapter 2 explains the models and methods used in this work, Chapter 3 discusses the different machine learning algorithm, while Chapter 4 provides a review of previous works. The results and conclusions are presented in Chapter 5 and Chapter 6, respectively.

2 MODELS AND METHODS

In this chapter, I will present a general idea of the Ising model since this model covers a wide field of study. In addition, I will give specific concepts about it, which will help the reader to understand this thesis, by only focusing on the systems studied in this work.

2.1 Ising model

The Ising model is one of the simplest but non-trivial model of interacting spins. The Ising model in one dimension was proposed by Lenz in 1920, but later in 1925 was discussed by Ernst Ising [2.1]. The exact solution of the two-dimensional Ising model was made by Onsager in 1994 [2.2].

The Ising model represents a magnet as a lattice in which each site of the lattice has a magnetic moment represented by a spin. The lattice can be in any dimension and any lattice structure. In our case, we studied the square and the hexagonal lattice. The spins will take the values of 1 for spins up and -1 for spins down. We can study the evolution of the system over time depending on two variables, which are the interaction strength and temperature. The energy of the Ising model includes two contributions: the interaction between neighbor spins *J,* which induce a parallel alignment of neighbors (ferromagnetic or antiferromagnetic interaction), and the effect of an applied magnetic field on each spin. The energy of the square system is

$$H = -J \sum_{<i,j>} S_i S_j - h \sum_i S_i \qquad (2\text{-}1)$$

Where $< i, j >$ means the sum over the nearest-neighbor pair of spins. This means the spin at site i, j interacts with i $(j \pm 1)$ and $(i \pm 1)$ j assuming a periodic boundary condition; in other words, each spin interact with four other spins (see Fig 2-1, e.g. square lattice) regardless of their position on the finite lattice. J is the interaction, $J > 0$ for ferromagnetic model and $J < 0$ antiferromagnetic model. h is the external magnetic field which in our study will always be zero.

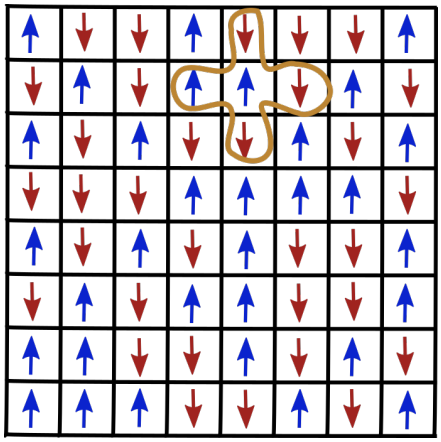


Figure 2-1. Ising Model spins representation in a square lattice

For this system Onsager [2.2] found the exact solution of the ferromagnetic-paramagnetic second order phase transition at

$$T_c = 2J/K_B \ln(\sqrt{2} + 1) = 2.2692(J/K_B). \qquad (2\text{-}2)$$

The Hamiltonian of the hexagonal ferromagnetic Ising model is the same as the square lattice system, but the difference is in the spin interactions, which only occur among the nearest neighbors. In the square lattice each spin interacts with four spins as is shown in Figure 2-1. For the hexagonal lattice the interaction is among three spins as shown in Figure 2-2.

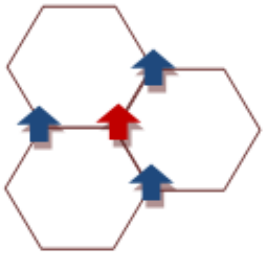


Figure 2-2. Schematics for the spin interactions of the hexagonal lattice.

For the hexagonal lattice, a phase transition to an ordered state occurs when the temperature T is reduced below the critical value [2-3],

$$T_c = 2J/K_B \ln(2+\sqrt{3}) = 1.519(J/K_B). \quad (2\text{-}3)$$

2.2 Phase transition and finite size scaling

The phase transition is one of the most common topics studied in physics in many problems such as crystal melting, ferromagnetism, Ising model among others, and it has

required the use of simulation for further study. The phase transition can be described by an **order parameter**, which is some thermodynamic quantity that becomes nonzero below the critical temperature and is used as a measure of the transition into an ordered phase. Mathematically, the order parameter is zero in one phase. Normally this is the disordered phase and the non-zero phase is the ordered phase. There are several transition orders, but most common are the 1st and 2nd order, which refer to the number of derivatives of the free energy.

In the case of the Ising model the 1st derivative of the free energy gives the magnetization *M* and the 2nd derivative gives the magnetic susceptibility χ_m

$$M = \frac{1}{V}(dF/dH)_{h=0} \tag{2-4}$$

$$\chi = \frac{1}{V}(d^2F/dH^2)_{h=0} \tag{2-5}$$

The order transitions are classified by their critical exponents, which characterize the behavior at the critical point and establish universality classes for phase transition [2.5]. The most important are:

Magnetization. $M \sim |T - T_c|^{\beta}$

Magnetic susceptibility. $\chi \sim |T - T_c|^{-\gamma}$

Heat capacity. $C \sim |T - T_c|^{-\alpha}$

Correlation length. $\xi \sim |T - T_c|^{-\nu}$

The critical exponents for the 2D Ising model are known exactly [2.5]:

$$\beta = 0.125 \qquad \alpha = 0 \qquad \gamma = 1.75 \qquad \nu = 1$$

On a finite lattice, which is our case, we have a finite number of degrees of freedom and everything is analytic, presenting no divergences. In an infinite system (L∼∞), the correlation length (domain size) diverges or becomes infinite at T_C. However, as we have a simulation in finite size, when the correlation length is $\xi \sim L$ the system has already become ordered. Then we can say that the system has a pseudocritical point when:

$$\left|T_{c(\infty)} - T_{c(L)}\right|^{-\nu} \sim L. \tag{2-6}$$

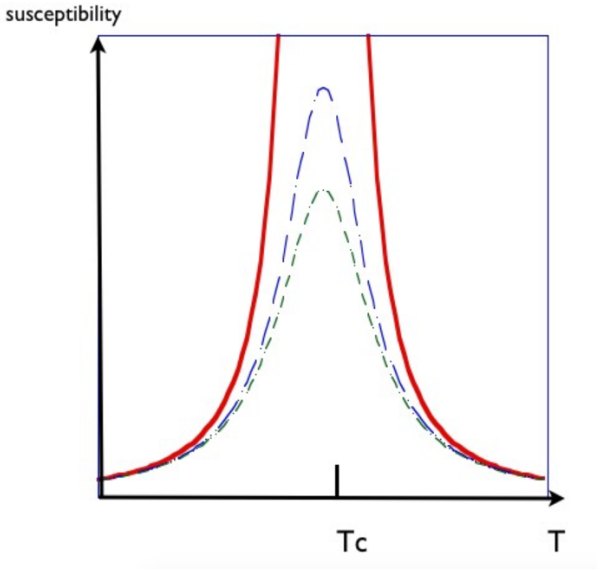


Figure 2-3. Susceptibility versus temperature for finite (dash lines) and infinity (red line) systems taken from [2.6].

If we consider the susceptibility for an infinite system, it becomes infinite at the critical temperature, but for finite systems we can say that the susceptibility χ has a maximum at $T_{c(L)}$. The susceptibility at that maximum point is

$$\chi \sim L^{\gamma/\nu}. \tag{2-7}$$

With the analysis mentioned before about the finite size scaling, we can determine the critical temperature $T_{c(\infty)}$, using various lattice sizes and locate the maximum of χ. Also, we can to determine the critical exponents ν and γ making a power law fit to the maximum location of χ.

$$T_{c(L)} = T_{c(\infty)} - aL^{x}\ , \tag{2-8}$$

where $T_{c(L)}$ is the critical temperature of the different lattice sizes taken from the maximum of the susceptibility, a is a constant to be found with the fit, and x will be $-1/\nu$ after the fit. For an infinite system the order parameters are always zero or non-zero. For the ferromagnetic Ising model it is the total magnetization per site

$$M = \langle \frac{1}{N} \sum_{i} s_i \rangle\ . \tag{2-9}$$

As we work with a finite size, the order parameter becomes, on overage, zero for all temperatures. Of these, $\langle |M| \rangle$ is usable, since it is almost zero in the symmetric phase:

$$M = \langle \left| \frac{1}{N} \sum_{i=1}^{N} s_i \right| \rangle. \tag{2-10}$$

2.3 Monte Carlo Method

The Monte Carlo simulation helped us to generate the different configurations to feed the Machine Learning algorithm of this work. This method is used to calculate numerically the thermodynamic properties, as averages in a system using numerical simulations. The idea of the method is to find an algorithm to generate a long sequence of configurations of a system, such that after a while each configuration is generated with the adequate probability to describe the equilibrium of the system. For example, to simulate a system at a constant temperature T, each configuration X must be generated with the probability (frequency). The purpose of performing a Monte Carlo simulation is the generation of an appropriate random set of states according to the Boltzmann probability distribution. More specifically, the probability p_μ that a system is in a given state μ with energy E_μ is:

$$p_\mu = \frac{e^{-\beta E_\mu}}{Z} \tag{2-11}$$

Where $\beta = 1/K_B T$ and Z is the partition function:

$$Z = \sum_{\mu} e^{-\beta E_\mu} \tag{2-12}$$

We are using in Markov processes, which describe how given a system in one state μ, a new state of that system ν is reached [2.7]. The probability of generating the state ν given μ is called the transition probability $P(\mu \to \nu)$ and should satisfy the conditions: 1) they should not vary over time, 2) they should depend only on the properties of the current state μ and ν, and 3)

$$\sum_{\nu} P(\mu \to \nu) = 1. \tag{2-13}$$

A Markov process must also fulfill the condition of detailed balance. That is the equilibrium condition of the Boltzmann probability distribution, which establishes that the rate at which the system makes transitions into and out of any state μ must be equal.

$$\sum_{\nu} P_{\mu} P(\mu \to \nu) = \sum_{\nu} P_{\nu} P(\nu \to \mu). \tag{2-14}$$

Using the equation 2-13

$$P_{\mu} = \sum_{\nu} P_{\nu} P(\nu \to \mu) \tag{2-15}$$

An additional condition to our transition probabilities is:

$$P_{\mu} P(\mu \to \nu) = P_{\nu} P(\nu \to \mu), \tag{2-16}$$

Thus, preventing the system from entering what is called a limit cycle, which means it can enter in a trajectory that is a closed cycle of states [2.8].

If we satisfied the conditions mentioned before, the equilibrium distribution of state in Markov process will be the Boltzmann distribution

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{P_\nu}{P_\mu} = e^{-\beta(E_\nu - E_\mu)}. \tag{2-17}$$

2.3.1 The Metropolis Algorithm

The Metropolis algorithm was made by Nicolas Metropolis and co-workers in 1953 [2.9]. The idea of the Metropolis algorithm is to choose a set of selection probabilities, one for each possible transition from one state to another, $\mu \to \nu$, and then choose a set of acceptance probabilities $A(\mu \to \nu)$. The algorithm works by repeatedly choosing a new state ν, and then accepting or rejecting it at random with the chosen acceptance probability. If the state is accepted, the system changes to the new state ν. If not, it just leaves it as it is, and the process is repeated again and again.

In this algorithm the selection probability $g(\mu \to \nu)$ for each of the possible states ν are all chosen to be equal

$$g(\mu \to \nu) = \frac{1}{N} \tag{2-18}$$

Where N , is the total number of states. The condition of detailed balance takes the form

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{g(\mu \to \nu)A(\mu \to \nu)}{g(\nu \to \mu)A(\nu \to \mu)} = e^{-\beta(E_\nu - E_\mu)} \tag{2-19}$$

One possible acceptance ratio that satisfices this equation is

$$A(\mu \to \nu) = A_0 e^{-\frac{1}{2}\beta(E_\nu - E_\mu)} \tag{2-20}$$

The constant of proportionality A_0 cancels in 2-19, so we can choose any value for it, but as is a probability, the maximum value that it can takes is 1. For a better understanding, suppose that the state μ has the lower energy, and the state ν has the higher $E_\mu < E_\nu$. Based in this, the larger acceptance ratio $A(\nu \to \mu)$ is set to one, and in order to satisfy the equation 2-19, $A(\mu \to \nu)$ must take the value $e^{-\beta(E_\nu - E_\mu)}$, in other words we can say:

$$A(\mu \to \nu) = \begin{cases} e^{-\beta(E_\nu - E_\mu)}, & if\ E_\nu - E_\mu > 0 \\ 1, & otherwise. \end{cases} \tag{2-21}$$

2.3.2 Implementation of the Metropolis Algorithm in the Ising Model

For the implementation of the Metropolis algorithm in the Ising model, we considered the case of zero magnetic field and a lattice of spins to work with. We used periodic boundary conditions to ensure that all spin have the same number of neighbors and local geometry.

Based on the theoretical critical points of the square and the hexagonal lattices, we chose our temperature range in which we worked. We started with a disordered random state from a temperature above the critical point and ended at the equilibrium temperature. To start the simulation, we just pick a random single spin k to be flipped. Next, we calculate de difference in energy before and after the flip $E_\nu - E_\mu$, calling ν the new state and μ the old one. The way to calculate this energy difference is by substituting the values s_i^μ of spins in

state μ into the Hamiltonian to calculate E_μ , then flip the spin chosen k, and calculate E_ν. The change in energy between the two state is thus

$$E_\nu - E_\mu = -J \sum_{\langle ij \rangle} s_i^\nu \, s_j^\nu + J \sum_{\langle ij \rangle} s_i^\mu \, s_j^\mu \tag{2-22}$$

As we flip a single spin, most of the terms in the calculation in the energy difference don't change and the difference of energy is reduced to

$$E_\nu - E_\mu = -J \sum_{i \; n.n \; to \; k} s_i^\mu (s_k^\nu - s_k^\mu). \tag{2-23}$$

The sum is over the nearest neighbors of the flipped spin k and the fact that all these spins do not themselves flip, means that $s_i^\nu = s_j^\nu$. Now, if the spin that we chose is $s_k^\mu = +1$, then after it has been flipped we have $s_k^\nu = -1$, then $s_k^\nu - s_k^\mu = -2$ or in the other case that $s_k^\mu = -1$ and $s_k^\nu = +1$, $s_k^\nu - s_k^\mu = +2$. Based on this, we can write the difference of energy as

$$s_k^\nu - s_k^\mu = -2 s_k^\mu \tag{2-24}$$

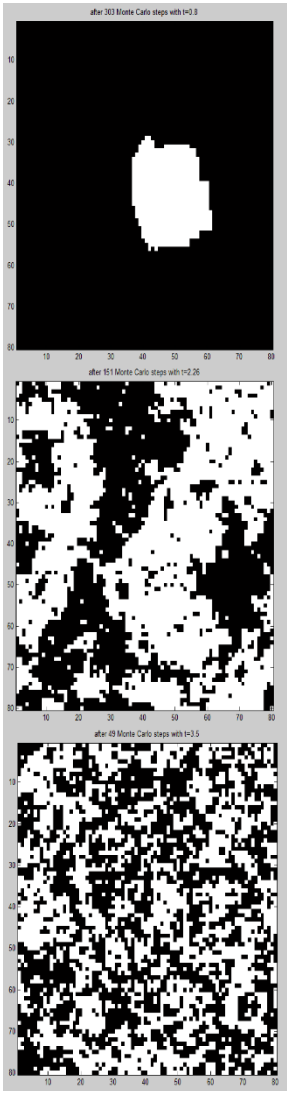
And so

$$E_\nu - E_\mu = 2Js_k^\mu \sum_{i\ n.n\ to\ k} s_i^\mu \tag{2-25}$$

Knowing the previous analysis, we can now implement the algorithm as follows:

- If $E_\nu - E_\mu \leq 0$ we definitely accept the move and flip the spin $s_k \to -s_k$.
- If $E_\nu - E_\mu > 0$ we still want to flip the spin with probability $A(\mu \to \nu) = e^{-\beta(E_\nu - E_\mu)}$.

In the algorithm evaluate the acceptance ratio $A(\mu \to \nu)$ by using the value of equation 2-25, and then we choose a random number r between zero and one. If the number r is less that our acceptance ratio $r < A(\mu \to \nu)$, then we flip the spin. If it isn't, we leave the spin alone. This process is repeated over and over, choosing a spin, calculating the energy change to see if we flip it, and then deciding whether to flip it according to the acceptance ratio.



$T < Tc$

$T \approx Tc$

$T > Tc$

figure 2-4. Monte Carlo simulation of the ferromagnetic Ising model below Tc, near Tc, and above Tc, for a finite-size scale. The black dots are the spins -1 and the whites are spins +1.

3 MACHINE LEARNING ALGORITHMS

For this work, we used supervised and unsupervised machine learning. In supervised machine learning, we have a dataset and already know what output to expect labelling the data, while unsupervised machine learning infers patterns from a dataset without reference to known outcomes. We can make predictions about a certain problem without knowing the labels by clustering the data, based on the relationships among the variables in the data [3.1].

In this work, two machine learning techniques will be implemented: Principal component analysis (PCA), an unsupervised technique, as well as Neural Networks (NN), a supervised technique.

3.1 Principal Component Analysis (PCA)

PCA is a powerful tool for compressing your data into small dimensions without much loss of information [3.2]. To understand this method, suppose that we have two variables x_1 and x_2, and a cloud of points. Each point $\boldsymbol{i}$ is represented with the variables as x_{i1} and x_{i2} as shown in Figure 3-1.

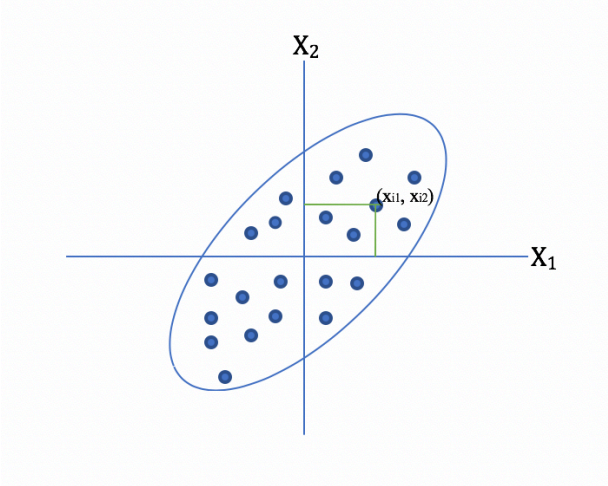


Figure 3-1. Representation of data in the variables.

We need to find the variable combination that collects the most information, this is the space direction in which we get the highest variance.

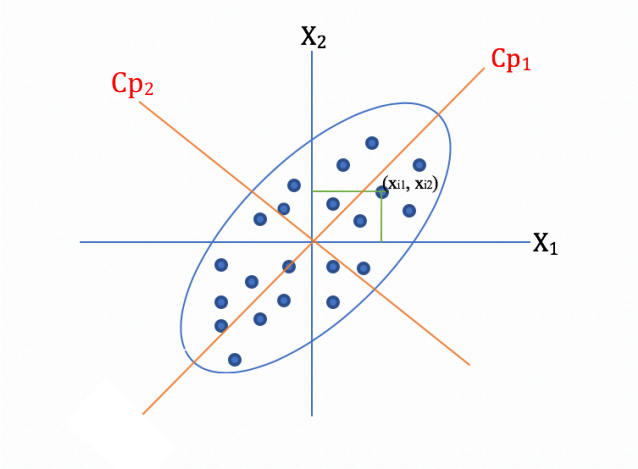


Figure 3-2. Principal component in the data.

It is expected that these directions collect the most information about the data. Once this direction has been found, we can see the data in terms of the principal components. Now, each point can be represented in terms of the principal component coordinates, y_{i1} and y_{i2}

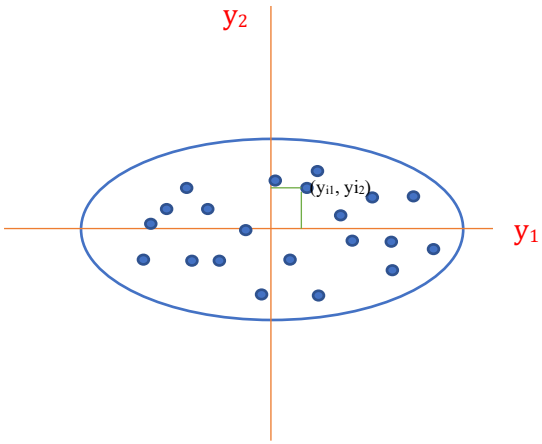


Figure 3-3. Principal component coordinates.

We can define the coordinate components in terms of the initial variables as:

$$y_{i1} = x_{i1}w_{11} + x_{i2}w_{21} \quad (3\text{-}1)$$

$$y_{i2} = x_{i1}w_{12} + x_{i2}w_{22} \quad (3\text{-}2)$$

where w_{ij} are coefficients called weights, which are vectors in the direction of those components.

In general, for N dimensions we have the following expression for the principal components:

$$\boldsymbol{Y_j} = X_1 W_{1j} + X_2 W_{2j} \ldots X_N W_{Nj} \tag{3-3}$$

And the coordinate component i in terms of the component j is:

$$y_{ij} = x_{i1} w_{1j} + x_{i2} W_{2j} \ldots x_{iN} w_{Nj} \tag{3-4}$$

For all the coordinate components we have

$$\begin{pmatrix} y_{1j} \\ . \\ . \\ . \\ y_{1N} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nN} \end{pmatrix} \begin{pmatrix} w_{1j} \\ . \\ . \\ . \\ w_{Nj} \end{pmatrix} \tag{3-5}$$

In matrix notation,

$$\boldsymbol{Y}_{\boldsymbol{j}} = \boldsymbol{X W_j} \tag{3-6}$$

for all the data set of components

$$\begin{pmatrix} y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nN} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nN} \end{pmatrix} \begin{pmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{XW} \tag{3-7}$$

3.1.1 Principal Components in the Ising Model

Data was collected by a Monte Carlo simulation in matrix S, with dimensions MxN, where $M = nT$, T is the number of different temperatures, and n is the number of configurations under the same temperature. Each row of the matrix will be a configuration sample.

$$S = \begin{pmatrix} 1 & 1 & -1 & \cdots & -1 & 1 & -1 \\ & \vdots & & \ddots & & \vdots & \\ -1 & 1 & -1 & \cdots & 1 & 1 & -1 \end{pmatrix}_{MxN} \tag{3-8}$$

Once we have matrix S, the next step is to apply PCA. S must be centered, subtracting the mean value $m_{ij} = (1/M)\sum_i S_{ij}$ of each column and the values from the entries in the column to obtain the matrix $\boldsymbol{X}$.

The PCA finds the principal components through a transformation vector of the original data:

$$Y = \boldsymbol{X}W \tag{3-9}$$

For this case, PCA will find or identify patterns in the data with temperature change. The main goal of PCA is to find out one or a few directions where you can group all the datasets with the fewest possible losses of information. This is equivalent to searching for the maximum variance in which the data is distributed.

The orthogonal transformation is due to vector $W = (w_1; w_2; \ldots; w_N)$, where $w's$ are called weights; the first weight is found by:

$$w_1 = \arg max_{\|w\|=1} \left\{ \sum_i (x_i . w)^2 \right\} \tag{3-10}$$

Instead of employing the maximum variance, the eigenvector corresponding to the largest few eigenvalues of the matrix $\boldsymbol{X^T}(covariance)$ can be found by:

$$X^T X w_n = \lambda_n w_n \tag{3-11}$$

The principal components are calculated as:

$$Y_n = X w_n \tag{3-12}$$

where $\boldsymbol{w_1}$ will be the vector corresponding to the largest variance, namely, the larger eigenvalue.

Results are based on the 'quantified principal components' that are defined as the average:

$$\langle |y_n| \rangle = 1/n \sum_n |y_n|. \tag{3-13}$$

3.2 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) have recently been widely used in machine learning. This is a technique inspired by the biological neurons in our brain and how they work. The neural network consists of a processing unit of neurons which have a direct weighted connection between other neurons [3-1]. The weights are responsible for the transfer of data between all neurons [3-1]. Like the neurons in our brain, the neuron of a neural network is represented in Figure 3-4, where the orange circle is the body of the neuron, the blue circles are the input and the gray one is the output.

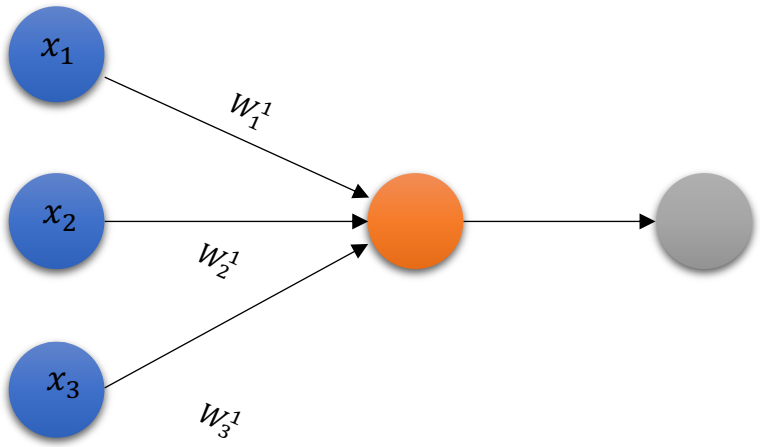


Figure 3-4. Neural neuron representation.

When we talk about neural networks, we are talking about many neurons connected to each other. An artificial neural network is composed of an input layer, hidden layers and an output layer. An example of this is seen in figure 3-5, where a neuron n_1 is connected with all the inputs that the other neurons are also connected to. n_1 could also be connected to many other neurons, creating other layers in the network or it could also be connected to the output of the network.

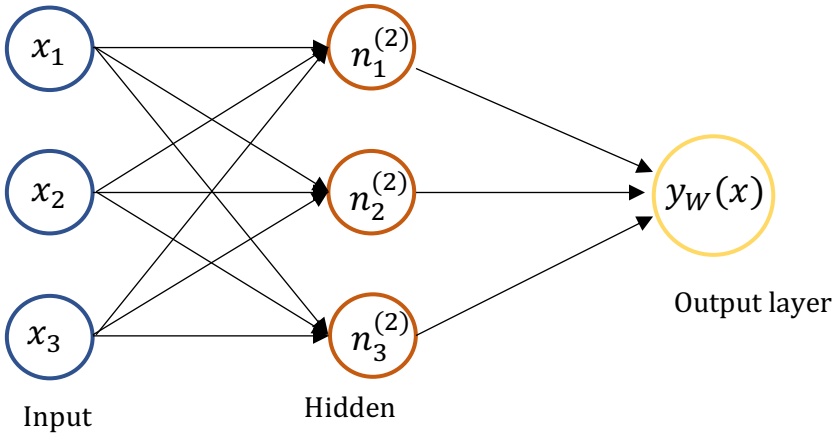


Figure 3-5. Simple Neural Network representation.

In Figure 3-5, a neuron also called perceptron, receives the inputs (x_1, x_2, x_3) and passes it though what is called an ***activation function.*** This calculates a weighted sum of its input, adds a bias and then decides whether it should activate the neuron or not. There are several activation functions, but for this work, we will only focus in the Sigmoid activation function, since the probability exist only between the range of 0 and 1.

$$F(z) = \frac{1}{1 + e^{-z}} \tag{3-14}$$

Where $z = \sum_j w_j x_j + b$, and b is the bias.

To understand what happens mathematically in a neural network, consider the ANN example in Figure 3.6.

Note that:

$n_i^{(j)}$ is the unit i in layer j

$W^{(j)}$ is matrix of weights controlling function mapping from layer j to layer $j + 1$ [3.1].

$$n_1^2 = F\Big(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b \Big) \tag{3-15}$$

$$n_2^2 = F\Big(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b \Big) \tag{3-16}$$

$$n_3^2 = F\Big(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b \Big) \tag{3-17}$$

$$y_W(x) = F\left(W_{11}^{(2)} n_1^{(2)} + W_{12}^{(2)} n_2^{(2)} + W_{13}^{(2)} n_3^{(2)} + b \right) \tag{3-18}$$

where $y_W(x)$ is the final output in the neural network of Figure 3-6.

The ANN presented in this thesis is based on Carrasquilla [4.2] and Dongkyu Kim's previous works [4.8]. The input data are the different Ising configurations from Monte Carlo method, the same data that was used in PCA. The ANN was only made for the hexagonal lattice, which is the highlighted system of study in this work. To illustrate how the ANN works in the Ising model, two neurons in the hidden layer and a single output were used as a toy model, see Figure 3-7. The input layer values are determined by the raw spin configurations sampled by Monte Carlo simulation. After studing the toy model, we used 300 neurons in the hidden layer to determine the critical temperature of the hexagonal Ising model.

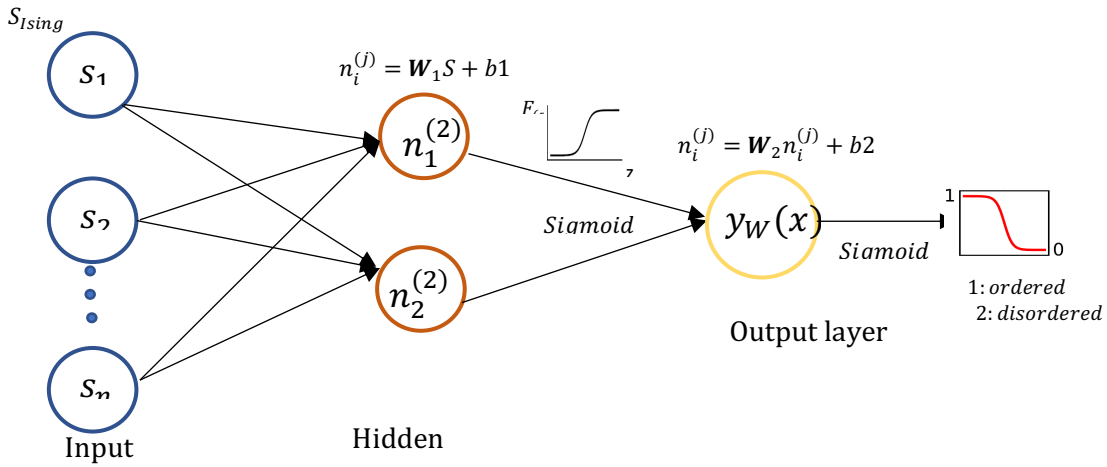


Figure 3-6. Network representation used in this work. Redrawn from [4.8].

3.3 Simulation details

An Ising model simulation was performed with the Monte Carlo method and implemented in Python. The Metropolis Algorithm was implemented for square and hexagonal ferromagnetic Ising models ($J > 0$) with zero magnetic field. The ferromagnetic interaction was set to $J = 1$. The details of the Monte Carlo simulation parameters are shown in table 3-1. 5000 configurations per temperature were collected, having a total of 545000 configurations for the square lattice and 620000 temperatures for the hexagonal system.

Table 3-1. Parameters in the simulation

Lattice	Lattice size (L)	M.C steps	T_{min}/T_{max} *(T/J)*	ΔT
Square	10,20,30,40,50	30000	0.8/3.5	0.025
Hexagonal	20,40,60,80	30000	0.82/2.08	0.01

The data collected from Monte Carlo was used to implement the machine learning techniques Principal Component Analysis and Artificial Neural Network. In NN, the data is divided in a training set and a testing set. In the training 196200 configurations were used, in the testing set 65400 and, in the validation, set 65400; for a total of 327000 configurations. 147000 of these configurations are below critical temperature and 180000 are above it. Configurations below this critical temperature were labeled 0 and the ones above the critical temperature were labeled 1. The network was trained on 100 epochs which is the number of

times that a given configuration passes through the neural network using the Adam optimizer with a learning rate of 0.001 to reduce the cross-entropy loss function. L2 loss was also applied to avoid overfitting with a $\lambda = 0.001$. This would give and average accuracy of 97%. The network was created with TensorFlow and trained on a 1080Ti GPU.

4 PREVIOUS WORK

Being able to classify and discover phase and phase transitions has been an important research topic in recent years and has been crucial in Condensed Matter Physics. In this chapter a thorough review of available literature on the discovery and the study of phases and phase transition with machine learning is presented. Since this study of machine learning is very recent, the literature will be presented in chronological order, highlighting the results and discoveries of each author.

In 2016, Lei Wang [4.1] explored the application of unsupervised learning to solve many-body physics problems that focus on phase transitions. They considered the Ising Model in a square lattice, for which 100 uncorrelated spin configuration samples were generated using Monte Carlo simulation at some preset temperatures. These configurations are fed to the unsupervised learning algorithm. To identify phase transitions, they used Principal Components Analysis (PCA) which is a widely-used feature extraction technique [4.1]. Principal components are in orthogonal directions in which the variances of the data decrease monotonically. PCA finds the principal components through a linear transformation of the original coordinates. When PCA is applied to the Ising configuration, it locates the most significant differences of the data changes versus temperature. These variations are relevant features in the data, and they can indicate a phase transition [4.1]. In their results, they found one dominant principal component as the temperature changed. The Ising configurations are most significant along the first PC. In this case, PCA has identified the order parameter of the Ising Model upon a phase transition. They projected the sample in the

space where the first two principal components are contained, showing the formation of two clusters, one associated with the high-temperature phase and the other one with the low-temperature phase. Once the baseline for applying the unsupervised learning technique is established in the prototypical Ising Model, they turn to a more complex case where the learner can no longer make any trivial findings. For this case, they considered the same Ising Model with a conserved order parameter (COP), where the occupation of each lattice site can be one or zero. In this case, they found four instead of one leading principal component. This indicated that in COP, the spin's spatial distribution changes considerably as the temperature varies [4.1]. They also show the structure factor versus temperature for different system sizes and demonstrate that the structure factor decreases as the temperature increases, an indicator of phase transition. The PCA finds out the structure factor related to symmetry breaking, which is crucial in phase transitions and condensed matter physics [4.1].

In the same year, Juan Carrasquilla and Roger G. Melko [4.2] identified phases and phase transitions using Feed-Forward Neural Network with TensorFlow (NN). They first employed the prototypical square-lattice ferromagnetic Ising Model, using Monte Carlo simulations to generate the spin configurations. They implemented the NN with an input layer by the spin configurations, a hidden layer of 100 neurons, and an analogous output layer. They used a cross-entropy cost function supplemented with an L2 regularization term to prevent overfitting. The NN was trained using the Adam method for stochastic optimization [4.3]. The neural network was able to identify the T_c with a 99% accuracy for L=40, indicating that this NN is capable of methodically closing in on the true thermodynamic value of T_c [4.2]. They used an already trained feed-forward neural network for the square lattice and applied it to a triangular lattice. The algorithm yielded a critical

temperature Tc/J =3.63581, which is close to the exact thermodynamic value (Tc/J = 3.640957), differing by less than 1%. Further, they applied these NN to problems of high interest in modern condensed matter, such as disordered or topological phases, in which they considered a two-dimensional *square ice* Hamiltonian. Through the neural network they were able to distinguish ground states from high-temperature states with 99% accuracy.

In 2017, Hu et al., [4.4] applied PCA to analyze the phase behavior and phase transition of various classical spin models, such as [4.4] square and triangular lattice Ising model, Biquadratic-exchange Spin-one Ising model (BSI), the Blume-Capel model, and the two dimensional XY model.

By applying the square lattice Ising model, Hu et al., were able to identify one dominant principal component. They plotted the measured first leading component versus temperature to mimic the magnetization of the system and also plotted the quantified second leading component to represent the susceptibility of the system. Then, they took the peaks from the graph of the quantified second-leading component and plotted them versus the inverse of the lattice dimension to obtain an approximate value of the critical temperature $T_c \sim 2.278 \pm 0.015$, which is close to the exact result $Tc/J \approx 2.269$ [4.5].

The order parameter of the system is the first principal component related with the first vector weigh W_1. They plotted the weight that corresponded to the second leading component (W_2) (see Figure 2-1a, taken from [4.4]), in this Figure, the second weight W_2 is compared with W_2' (see Figure 2-1b) which is a plot of:

$$W_2' = \frac{1}{L}[\cos(r_1 k_1), \ldots, \cos(r_N k_1)] + \frac{1}{L}[\cos(r_1 k_2), \ldots, \cos(r_N k_2)]. \qquad (4.1)$$

where r_i is the lattice site and $k_1 = (0, 2\pi/L)$, $k_2 = (2\pi/L, 0)$ are the lowest Fourier wave vectors. The first component was associated with the origin $k_0 = (0,0)$.

Based in the results mentioned above, they determined that for ferromagnetic Ising model in square lattice, PCA is building up in weight vectors which correspond to the Fourier modes of the spin configurations. finding an interesting result.

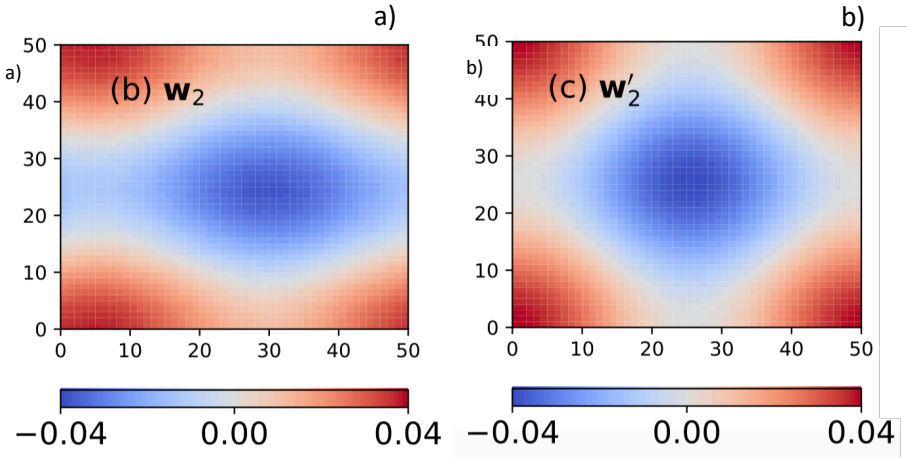


Figure 4-1 a) plotted weights of the second leading component, b) plot of the equation 2.1.

Also, in 2017, Ce Wang and Hui Zhai [4.6] studied phase and phase transitions working with Unsupervised Learning of Frustrated Classical Spin Models in triangular and union Jack lattices using the XY model for different temperatures. They also fed the algorithm with data generated by the classical Monte Carlo simulation and used Principal Components Analysis as a machine learning technique. They first analyzed the results based on a simple “toy model" in a square lattice which was useful for later discussion of triangular

and union jack lattices. PCA results showed two principal components which correspond to high and low temperature phases, that allowed the identification of a sphase transition at Tc = $0.9J^{29}$, below which a quasi-long range anti-ferromagnetic order is formed. Indeed, Wang and Zhai found that the two-major normalized eigenvalues are insensitive to system size for temperatures below Tc, and for temperatures above Tc. The two-major normalized eigenvalues decrease quickly as the system size increases.

For the triangular lattice, they performed a similar PCA analysis, collecting the data from Monte Carlo simulations at nine different temperatures. Results show four principal components with their corresponding eigen-vectors and revealed how the large principal components depend on temperature. Below a certain temperature, the four normalized major principle eigenvalues are not sensitive to system size, while above a certain temperature, they decrease to quite small ones as the system size increases. This transition temperature scale is consistent with the KT transition expected for the XY model in a triangular lattice [4.6]. For the union jack lattice, they found that four eigenvalues are considerably larger than the others which form the principal subspace. In this case, the low-temperature ordered phase has four sites at each unit cell. The PCA analysis shows four major eigenvalues at the lowest temperature and two which gradually vanish as temperature increases, similar to the square and triangular lattices.

Antonio Rebelo in 2017 [4.7] analyzed several physical models using Principal Component Analysis to study the square ferromagnetic Ising model, getting the same results of [4.4] and [4.1]. Next, he studied the both the square and triangular antiferromagnetic Ising models. For the square system, one leading principal component was obtained and again,

since the first principal component separates the configuration in clusters for low and high temperature, he plotted the second weight which corresponded to a Fourier mode of the system. In the triangular system, he obtained two dominant principal components, where the model remained disordered at every finite temperature with no obvious pattern that PCA could lock into. PCA, however, was able to distinguish between two phases when the data were projected in the first two principal components. After another treatment in the triangular Ising antiferromagnetic system, PCA yielded acceptable results for this frustrated model.

In 2018, Kim et al., [4.8] investigated the Ising model's phase transition learning. They found that having two hidden neurons with an optimized number of parameters through data-driven training was enough for an accurate prediction of critical temperature. They started observing the feed-forward network trained in the square lattice with a 50-neuron single hidden layer. They noticed that the weights between the input and the hidden layer were constant, meaning the input $\{S_i\}$ is reduced into its sum $\propto \pm \sum_i S_i$. They also observed neurons effectively unlinked with vanishing weights, suggesting that the hidden layer size could be smaller. Based on their observations above, they proposed a small network model containing only two neurons in the hidden layer. One neuron is linked with a positive constant weight $y \propto \sum_i S_i$, and the other neuron, with opposite sign $-y$. They showed that the two-neuron binary structure with two neurons successfully predicted the critical temperature in the different lattice systems, which aids in understanding the Ising model's supervised learning accuracy. They also found that the scaling dimension of the order parameter is embedded into the system-size dependence in the learning process.

5 RESULTS AND DISCUSSIONS

In this chapter, the PCA results for the study of the ferromagnetic Ising model in square lattices, for lattice sizes of L = 10, 20, 30, 40 and 50 are discussed, as well as the PCA results for the hexagonal lattices using lattice sizes of L = 20, 40, 60 and 80. Finally, results are presented for two neural networks that were performed for the hexagonal lattice system with the same lattice sizes used in PCA. The first neural network was built with two neurons in the hidden layer for system size L = 20, with the purpose of understanding the training and functionality effects of the neural network. Afterwards, a neural network with 300 neurons in the hidden layer was created to obtain an accurate approximation of the thermodynamic critical value T_c in the hexagonal system.

5.1 Square Lattice

Figure 5-1a presents the results of the first 10 principal components for the square system. It clearly shows one dominant principal component for the different lattice sizes. This means that in the Ising configurations, there is a dominant spin pattern; in this component, the Ising configurations vary significantly as the temperature changes [4.1], [4.4]. On the other hand, figure 5-1b shows a constant weight vector corresponding to this principal component. Notice that $w_1 \simeq 1/L$. This result demonstrates that with the first principal component we can identify the order parameter which corresponds to the

magnetization of the system, described by a given configuration from one of the rows of **X**. in equation (3-12).

$$Y_1 = \frac{1}{L}[s_1, s_2, s_3 \ldots] \begin{bmatrix} 1 \\ 1 \\ 1 \\ \ldots \end{bmatrix} = \frac{1}{L}\sum_{i=1}^{N} s_i \qquad (5\text{-}1)$$

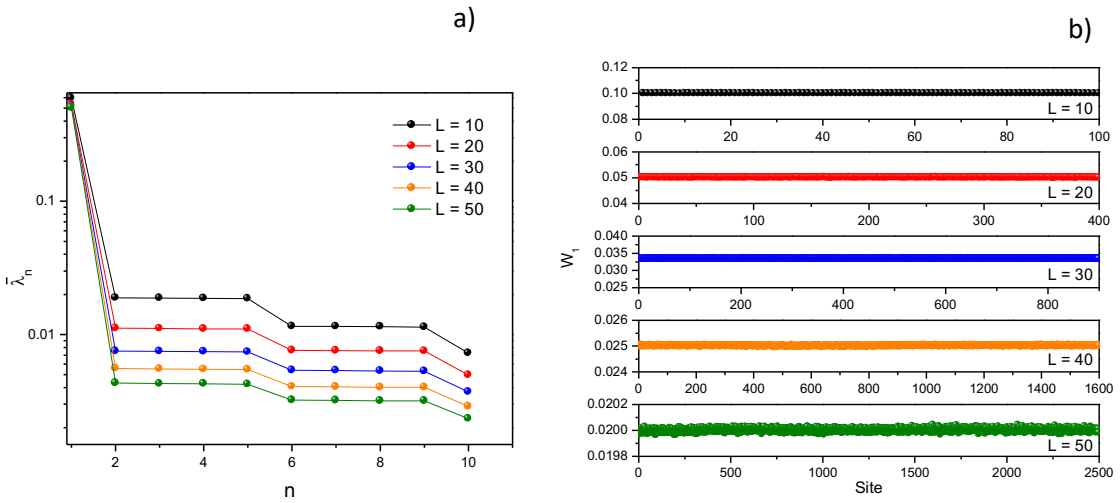


Figure 5-1. a) Shows the first ten PCA variance ratios from the Ising configurations for square lattice. b) Weights of the first principal component for each lattice size for square lattice.

A physical explanation of why PCA obtained a dominant eigenvalue in the ferromagnetic Ising model can be seen in the analysis made by Rebelo [4-7]. If a lattice with only two spin interactions is considered and s_1 is the spin of the first lattice site, the most favorable configurations of the ordered phase of the system would be of the form:

$$x_{ord} = (s_1, s_1)$$

and for the disordered phase:

$$x_{dis1} = (s_1, s_1) \qquad x_{dis2} = (s_1, -s_1)$$

When PCA is fed with configurations of both phases, the covariance [equation (3-1)] to be computed is:

$$C = X^T X = \frac{1}{M} \sum_n x_n^T \, x_n$$

where, M is the total number of configurations and x_n is the nth configuration.

If we consider a set of data p which corresponds to the ordered phase, then $(1 - p)$ will correspond to the disordered phase, hence, we can write:

$$C = pC_{ord} + (1 - p)C_{dis}$$

With enough data, it can be written:

$$C_{ord} = \langle x_{ord}^T x_{ord} \rangle$$

Where < > denotes the average over the value that x_{ord} can take, then for this case:

$$C_{ord} = \frac{1}{2} \sum_{s_1} s_1^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

For C_{dis}:

$$C_{dis} = \frac{1}{2} (\langle x_{dis1}^T x_{dis1} \rangle + \langle x_{dis2}^T x_{dis2} \rangle) = \mathrm{I}$$

Solving equation (3-1) for C_{ord}, yields the eigenvalues with their corresponding eigenvectors.

$$\lambda_1 = 2, \qquad w_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$$

$$\lambda_2 = 0, \qquad w_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix}$$

In the example considered, the results obtained in the PCA, were: a leading eigenvalue and a constant vector corresponding to this eigenvalue. For the disordered phase, the identity matrix was obtained, therefore, the greatest eigenvalue came from the ordered phase. This example can be generalized for a matrix of size LxL and for the case of the ordered phase, where only a matrix NxN of ones would be obtainable.

Solving for the eigenvalues and eigenvector of this matrix, we would obtain a characteristic polynomial of the form $(\lambda - N)\lambda^{N-1}$, where the highest eigenvalue would always be $\lambda_1 = N$ and its vector:

$$w_1 = \begin{bmatrix}1\\1\\.\\.\\.\\1\end{bmatrix}$$

The analysis done previously demonstrates in the simplest way how PCA works and how the result for the first component is obtained, where only one spin was flipping while the others remained constants. Actually, the second component would be related to this first state of excitation mentioned above, but in this model with only two spins, it would be nonsensical. The interesting aspect about PCA is that it detects this behavior for a full data set in a full lattice.

In Figure 5-2a, the quantified first leading components $<y_1>$, over the system size L versus temperature is plotted, in order to mimic the magnetization in the square system.

$$m = \frac{1}{N}\sum_{i=1}^{N} s_i = \frac{< P_1 >}{L} \tag{5-2}$$

where $\boldsymbol{m}$ is the magnetization and $N = L\ x\ L$.

By increasing the size of the system, the phase transition approaching the theoretical value represented with the dash line can be appreciated. Figure 5-2b, on the other hand, shows the plot of quantified second leading components, $<p_2>$, versus temperature, in this case, it represents the behavior of susceptibility (χ).

Finally, Figure 5-2c depicts the spin configurations in a two-dimension spanned plane, where three different clusters can be observed. The points close the origin corresponds to the spin configuration above the critical temperature T_c and the separate other two regions correspond to the spin configurations below this critical temperature. With these results, PCA shows that there is a phase transition in the 2D square Ising system.

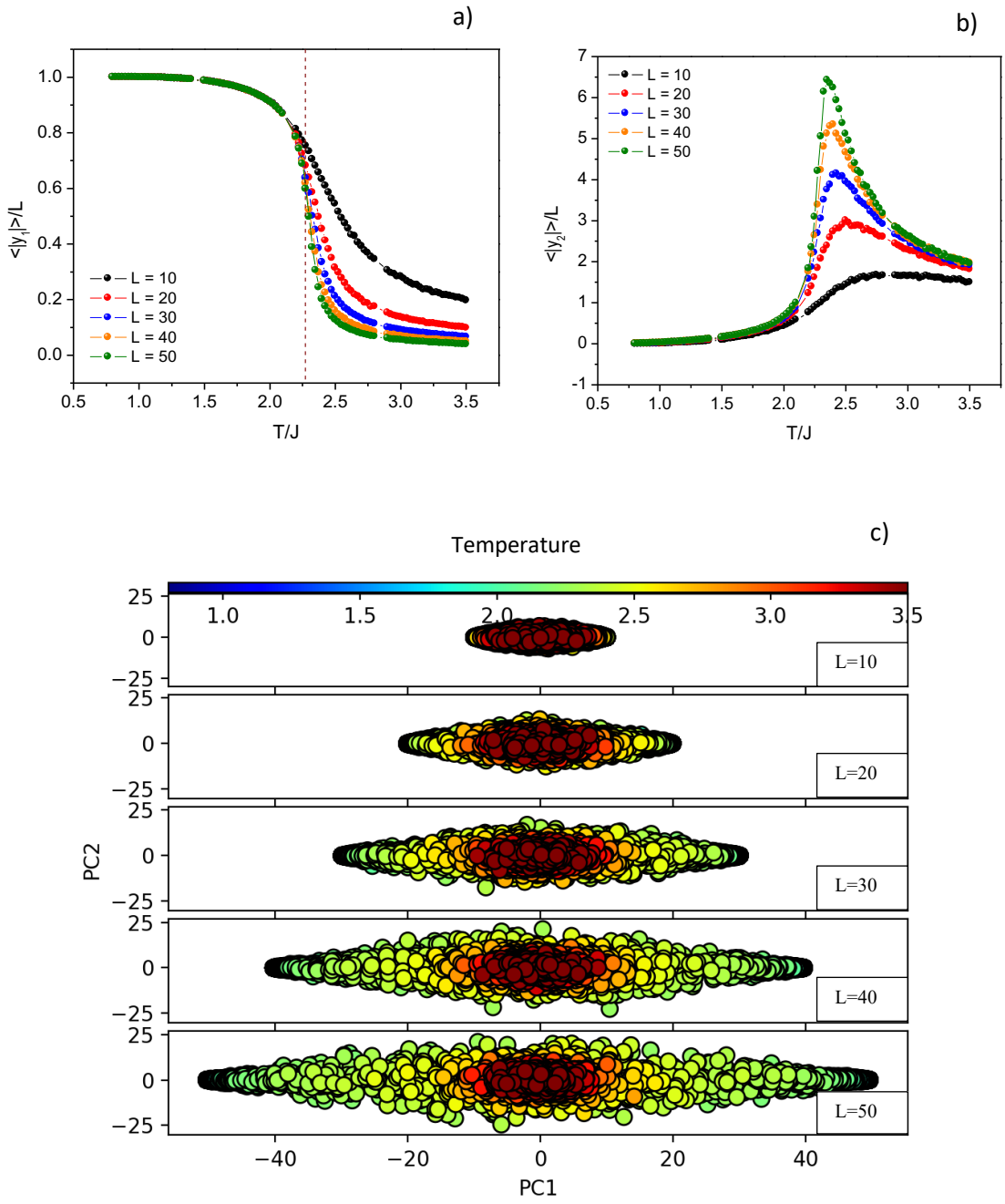


Figure 5-2. a) The normalized, quantified first leading component versus temperature which represents the magnetization of the system. b) The quantified second leading component versus temperature which represents the susceptibility of the system. c) Projection of the spin configurations onto the plane for the two principal components for a lattice of size 10, 20, 30, 40 with 50 configurations for each temperature.

Following the treatment in chapter 3, based on the finite scaling relation in thermodynamic limit:

$$T_c' = T_c + KL^{-1} \tag{5-3}$$

where T_c', is the maximum of the susceptibility (χ), T_c and K are the fitting parameters.

An approximation of the critical temperature T_c was obtained by plotting the maximums in Figure 5-2b versus the inverse of the system size, 1/L. The intercept of the fit gave an excellent approximation of T_c for the square system. We estimate the critical temperature to be $T_c = 2.26339\, J/K_B$, which is close to the exact thermodynamic $T_c = 2.2692 J/K_B$, differing from the true thermodynamic critical T_c by less than 0.5%. This result means that PCA was able to make a good prediction of the critical temperature T_c in the 2D Ising model (see Figure 5-3).

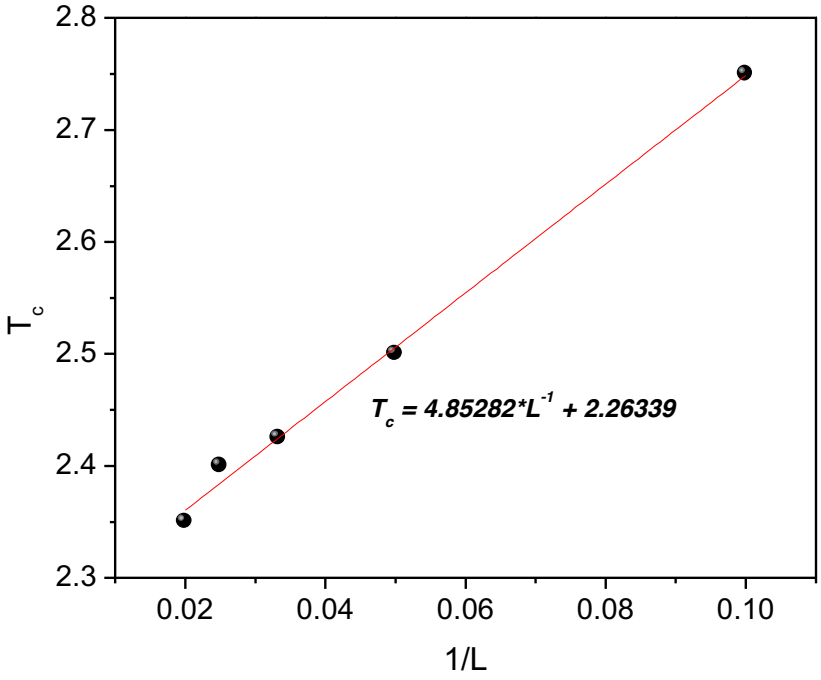


Figure 5-3. Critical temperatures taken from the maximums of Fig 5-2b versus the inverse of the lattice size.

Considering the weight vector of the second component (w_2) for the square lattice, a plot of the elements of w_2 was made using 7000 samples per temperature, (see Figure 5-4a). This plot is compared with the plot of the following equation represented in Figure 5-4b:

$$W_2' = \frac{1}{L}[\cos(r_1 k_1), \ldots, \cos(r_N k_1)] + \frac{1}{L}[\cos(r_1 k_2), \ldots, \cos(r_N k_2)] \qquad (5\text{-}4)$$

where r_i is the lattice site and $k_1 = (0, 2\pi/L)$, $k_2 = (2\pi/L, 0)$ are the lowest Fourier wave vectors. The first component is associated with the origin in $k_0 = (0,0)$.

It was determined in [4,4] that for the ferromagnetic Ising model in square lattice, PCA is building up in weight vectors which correspond to the Fourier modes of the spin configurations.

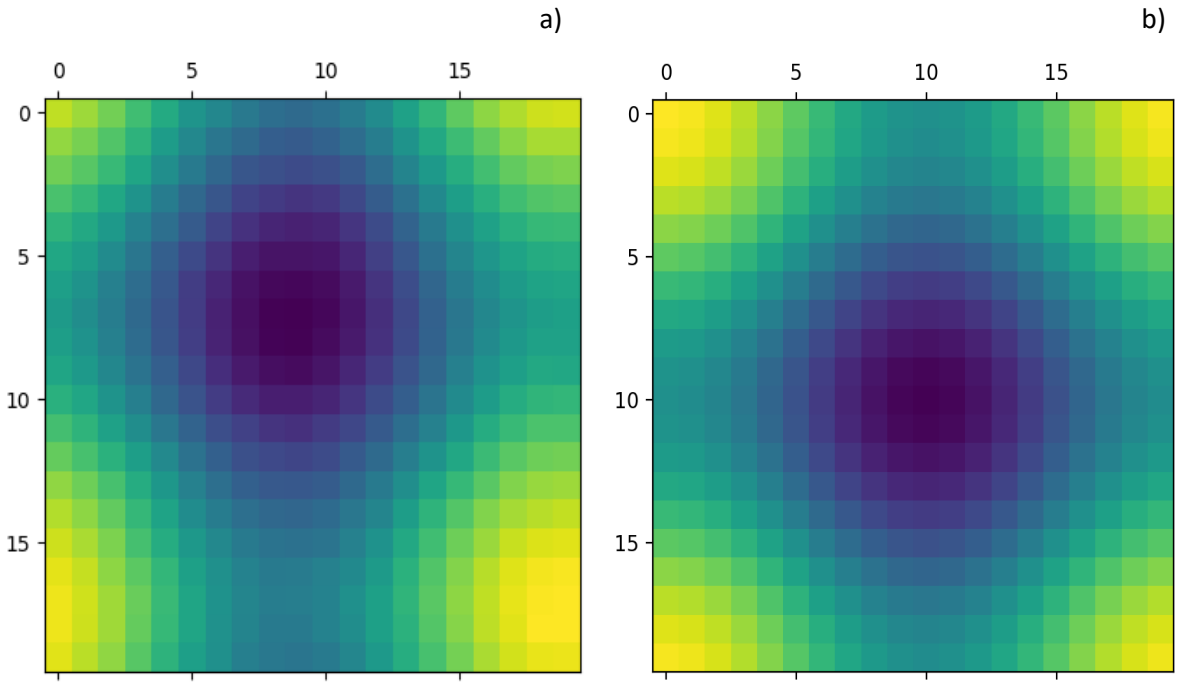


Figure 5-4. a) Weighs for the second component plotted on the square lattice L=20. b) plot of equation (5-4)

5.2 Hexagonal Lattice:

In the hexagonal system, we also noticed one leading principal component as in the square system, (see Figure 5-5a), meaning that in the hexagonal system, the data is confined in the first component. Figure 5-5b shows constant weights of the first principal component with values of $w_1 \simeq 1/L$ for each lattice size, again, the order parameter of this system was identified, namey magnetization.

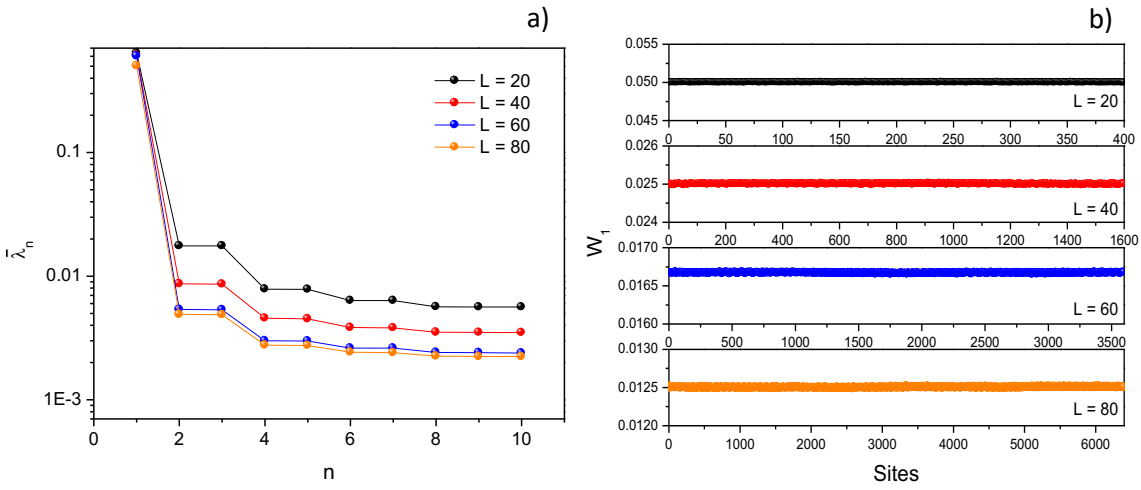


Figure 5-5. a) PCA first explained variance ratios from the Ising configurations for hexagonal lattice. b) Weights of the first principal component for each lattice size.

Figure 5-6a represents the magnetization of the hexagonal system. As can be seen, as the size increases, it gets closer to the teorical transition value. This makes sense due to the fact that the teorical critical value was calculated for finite lattices. The quantified second leading component versus temperature, shown in Fig 5-6b for the different lattice sizes (20, 40, 60 and 80), mimics the susceptibility in the hexagonal system. As discussed in Chapter 2, the susceptibility for an infinite system becomes infinite or diverges at the critical temperature. As we work with finite lattices, these have peaks, which run to the left, approaching the critical theoretical temperature. This behavior was also seen in the results for susceptibility in the square lattices.

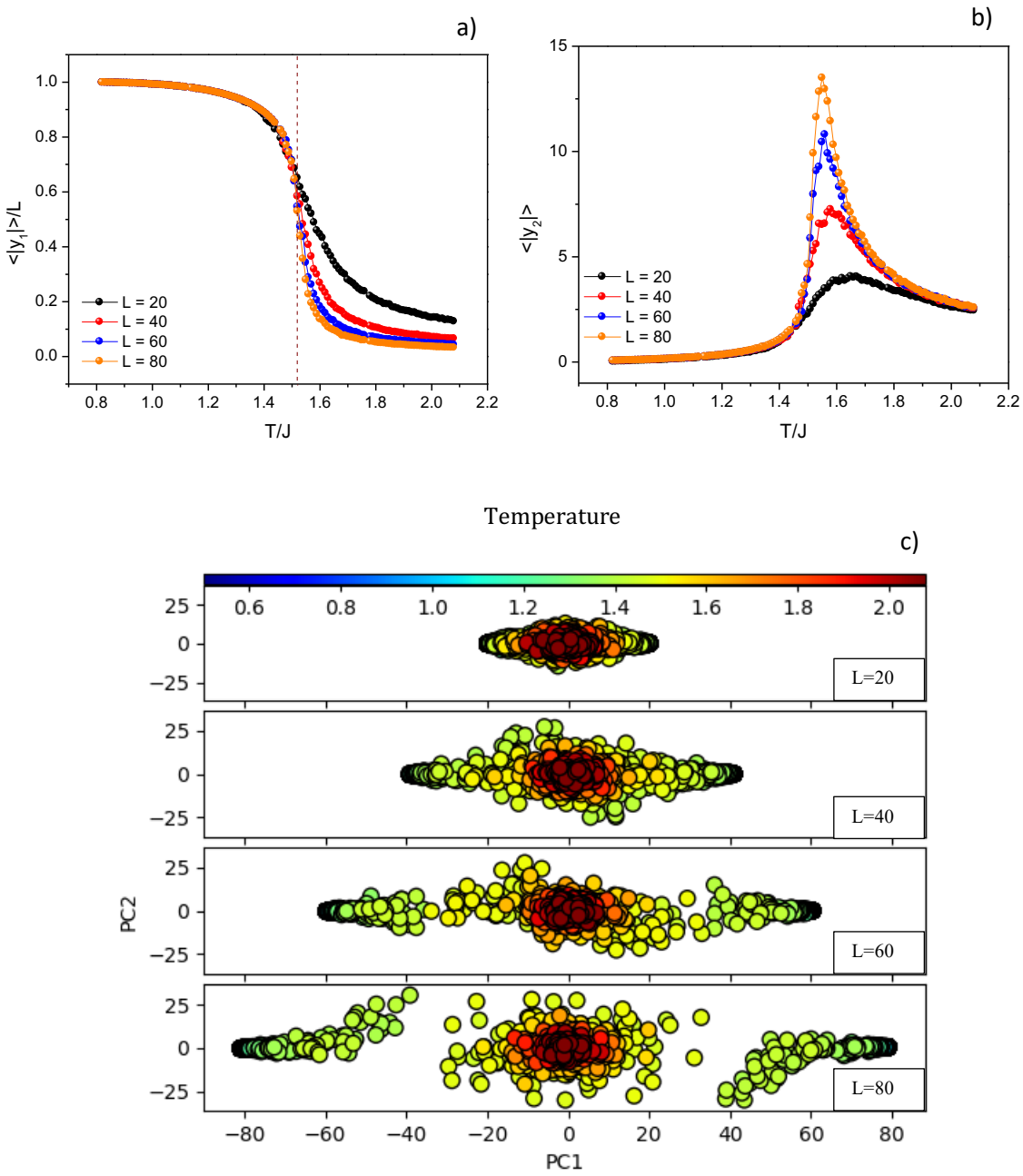


Figure 5-6. a) The normalized quantified first leading component versus temperature for hexagonal lattice which represents the magnetization of the system. b) The quantified second leading component versus temperature for hexagonal lattice which represents the susceptibility of the system. c) Projection of the spin configurations onto the plane for the two principal components for lattice of size 20, 40, 60, 80 with 100 configurations for each temperature.

Figure 5-6c depicts the projected spin configurations onto the two principal components plane for the different lattice sizes. Clearly, two clusters are also shown, this was also seen in the square lattice, where the points grouped in the center correspond to the highly disordered configurations, while group configurations in the extremes represent configurations in an ordered state, proving that a phase transition also occurs in the hexagonal lattice.

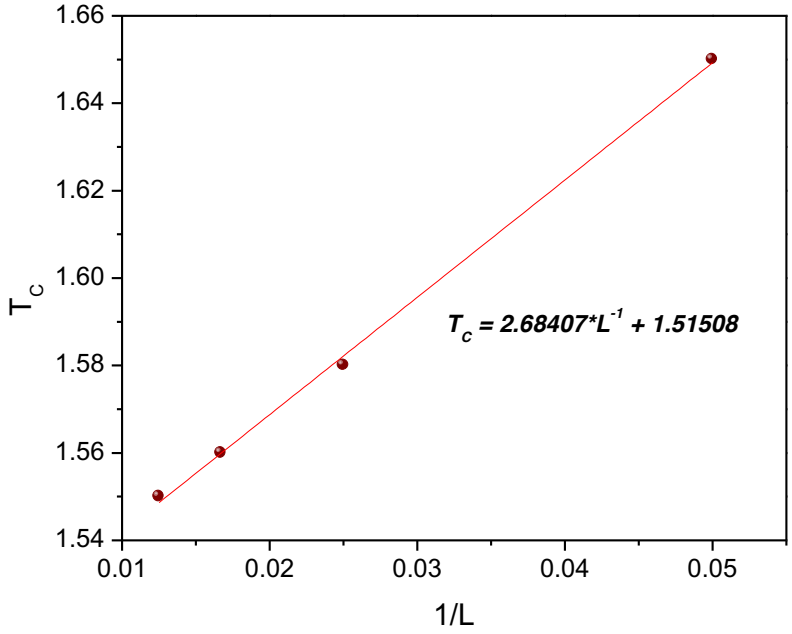


Figure 5-7. Critical temperatures taken from the maximums of Figure 5-6b versus the inverse of the lattice size.

In the hexagonal system, the finite scaling relation using equation (5-3) can be employed to determine the critical temperature T_c in the Ising hexagonal system ($T_c \approx 1.51508\, J/K_B$). This result is close to the exact theorical value in this system, which is $T_c \approx$

$1.519\, J/K_B$, representing a 0.5% percent error from the true thermodynamic critical temperature. This means that PCA is able to work with different lattice types.

The results of the weight for the second component plotted on the hexagonal lattice were different from the square lattice (see Figure 5-8a). Based on the analysis from [4.4], we could say the hexagonal lattice system studied in this thesis has different Fourier modes of the spin configuration. The Fourier mode corresponding to the second weight is defined by equation 5-5. The graph of this Fourier mode is seen in Figure 5.8b which is compared with the result given by PCA (Figure 5.8a)

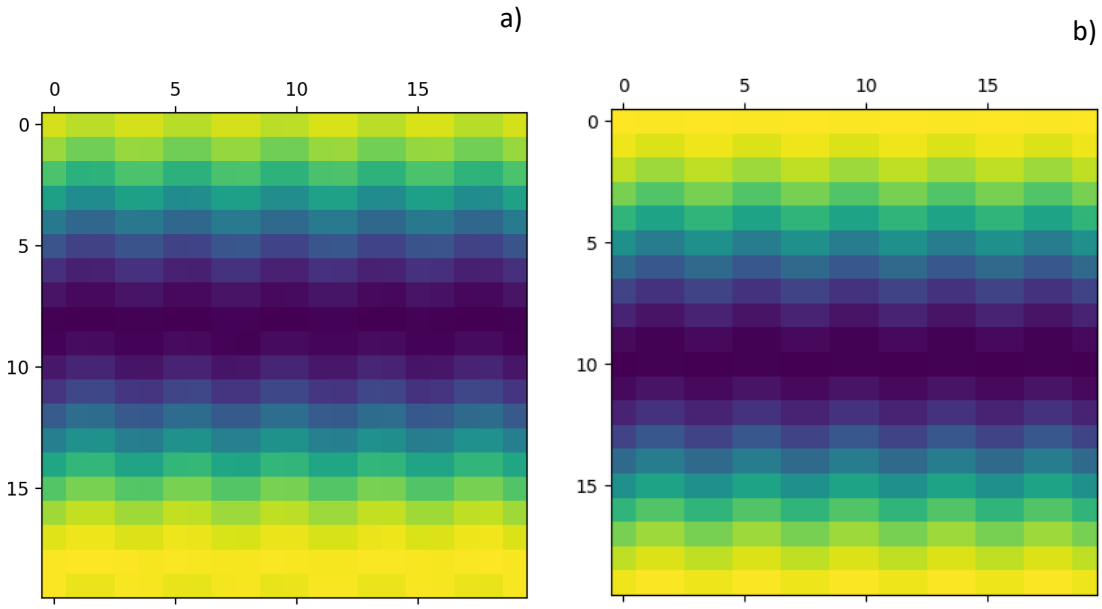


Figure 5-8. a) The weighs for the second component vector plotted on the hexagonal lattice L=20 b) plot of the equation (5-5)

$$W'_{2hex} = \cos\frac{1}{L}[\cos(r_1 k), \ldots, \cos(r_N k)] \tag{5-5}$$

where r_i is the hexagonal x-coordinate lattice site and $k = (2\pi/L, 0)$.

5.3 Neural Network Results

A test neural network was built with only two neurons in the hidden layer for a size L = 20 hexagonal system to explain the training effects of the parameters weights (W) and bias (b). Afterwards, a neural network with 300 neurons in the hidden layer was created to obtain an accurate approximation of the thermodynamic critical value T_c in the hexagonal system.

In Figure 5-9a, the argument of the neural network $W_o x + b_0$ for the input layer is displayed as a function of magnetization $m(x)$ of the configurations x. W_o and b_0 are the weights and bias at training iteration $t = 0$, which were randomly initialized. Once the training begins, the parameters are adjusted, having an accuracy of $\simeq 94\%$. The components of the vector $W_t x + b_t$ become approximately a linear function of the magnetization $m(x)$, as depicted in Figure 5-9b.

Figure 5-10 shows the weights adjusted after the training. In one neuron, the weights become positive constants while in the other neuron, the weights are constant but with opposite sign. This result gives an indication that the weights are responsible for magnetization.

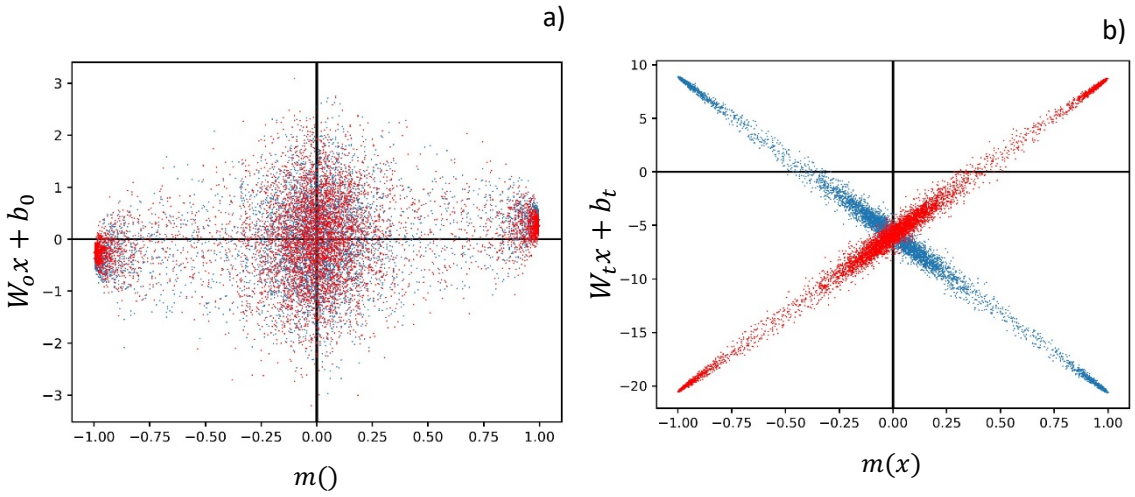


Figure 5-9. a) Arguments for a neural network with two sigmoid neurons in the hidden layer before the training. b) Arguments after the training.

The results shown in Figure 5-9 and Figure 5-10 explain that in the hidden layer of the neural network, the magnetization $m(x)$ is encoded and learned, corroborating Carrasquilla and Melko's research when using three neurons in the hidden layer [4-2]. These results also indicate that with only two neurons, the order parameter, $m(x)$, can be found, and a prediction of the critical temperature, T_c, of the system can be made, as seen in Figure 5-11.

It is necessary to remark that the neural network does not have prior knowledge for the search of the critical temperature and the order parameter of the system, the NN only has the different configurations made by the Monte Carlo simulation.

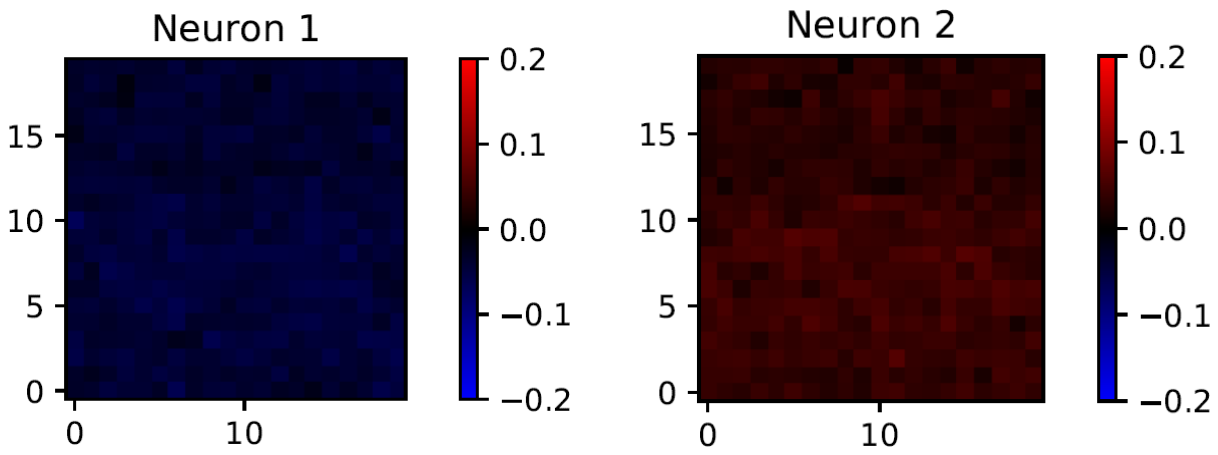


Figure 5-10. Weight values in each neuron after the training.

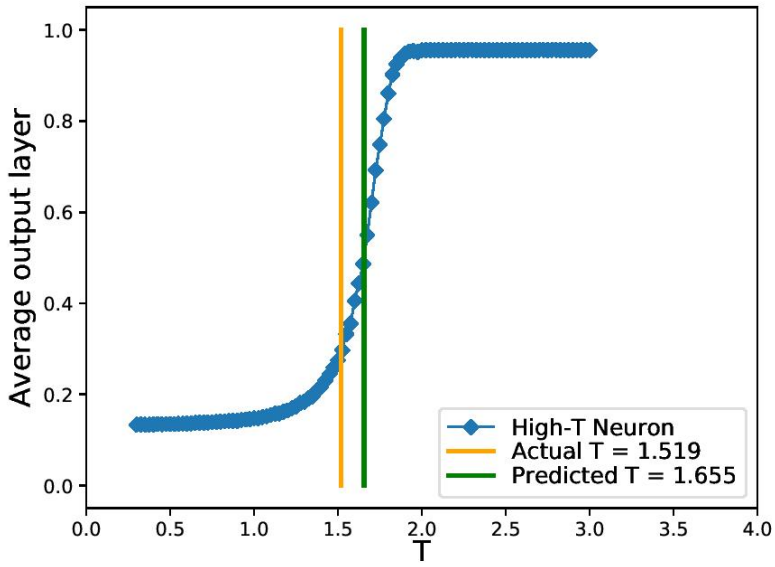


Figure 5-11. Average output layer versus temperature T of a NN with two neurons in the hidden layer for the hexagonal lattice of L=20.

To improve the accuracy between the predicted value and the theoretical value of the critical temperature Tc, a neural network with 300 neurons in the hidden layer was implemented for hexagonal lattice sizes of 20, 40, 60, and 80.

For L=20, the neural network was able to correctly classify 95 % of the uncorrelated data in the test set, at the same temperature as the training set. As the system increases, the classification of the uncorrelated data improves, obtaining a 97% of accuracy for system with L=80. The closer predicted temperature $T_c = 1.546 J/K_B$ was obtained for the system size L= 80, which make sense, since the real critical temperature of this system was calculated for an infinite system. Therefore, as the lattice size becomes larger, the transition temperature is closer to the real thermodynamic value. Also, by having a greater number of neurons in the hidden layer, the neural network learns better, given better accuracy results for the critical temperature.

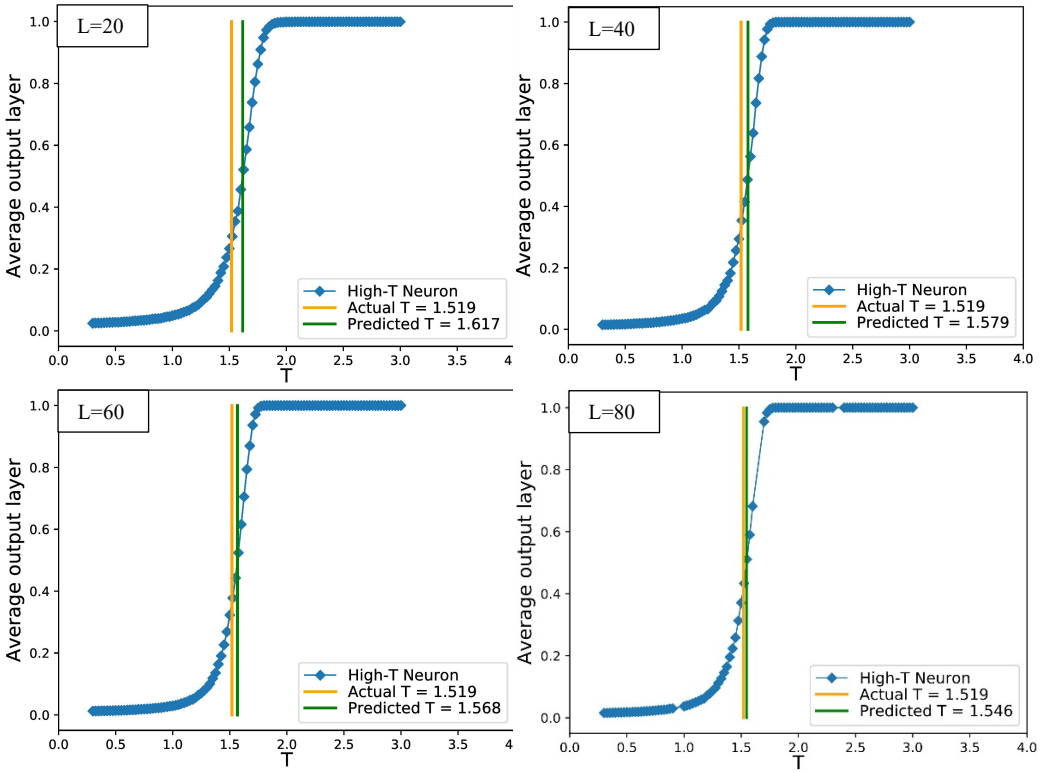


Figure 5-12. Average output layer versus temperature T of a NN with 300 neurons in the hidden for L=20, 40, 60 and 80 in hexagonal lattice.

The discrepancy of the results in both PCA and NN for the predicted temperature in square and hexagonal systems can easily be attributed to finite-size effects [4.2] and to the temperature resolution used in each system.

In PCA, the average of the second component y_2 had to be calculated for a given temperature value, which generates another error factor. Based on the above, it is impossible to obtain an error difference between de predicted value of the critical temperature and the theoretical value for both systems.

6 CONCLUSIONS

In this work, an unsupervised machine learning technique Principal Component Analysis was employed to study phase transitions in the Ising model for two specific systems: the square and hexagonal Ising model systems. The supervised machine learning technique Artificial Neural Network was also implemented for the hexagonal system.

Principal Component Analysis was able to recognize phase transitions in the Ising model. PCA results yielded a leading component and a constant weight vector, related to this particular component, for both systems. From this main component, the order parameter of the system was identified. It was possible to mimic the magnetization and the susceptibility in the Ising model with the two-leading component. Also, the critical temperature T_c in the square and hexagonal systems were determined. For the square system a $T_c = 2.26339\, J/K_B$ was obtained and for the hexagonal system a $T_c = 1.51508\, J/K_B$, having a 0.5% percent error from the true thermodynamic critical temperature.

When PCA was fed with spin configurations from Monte Carlo, spatial order patterns were recognized by clustering the data between the ordered and disordered phases as shown in Figures 5-2c and 5-6c. An interesting fact about the PCA technique for the ferromagnetic Ising model is that the weight vectors correspond to the Fourier modes of the spin configurations. The first weight vector is associated with the ordered phase and is enclosed in a single point, $k_0 = (0,0)$; hence, the physics of the Ising model is shown in a single dominant eigenvalue.

However, careful analysis reveals that further components also encase relevant information about the system. For example, information about the susceptibility of the Ising model for both the square and the hexagonal systems could be obtained from the second component. The weight vector associated with this component shows that the second Fourier mode corresponding to the spin's configurations remains an unknown that can be addressed by future research. The other Fourier modes can potentially explain the physical meaning of other missing components, which could contribute to the Ising ferromagnetic model.

It must be noticed that when employing PCA, no instruction is given to the model to look for physical properties, such as magnetization and susceptibility, nor to search for the Fourier modes of the different configurations. The model is only given the different configurations, which are provided from Monte Carlo simulation.

Regarding the results obtained though NN, we observed the performance of the neural network based on the Ising model. With only two neurons in the hidden layer, the effect of the neural network training could be seen. It was determined that the order parameter is encoded in the training effect in the hidden layer. The weights (W) were constant after training was achieved, observing the relationship they share with the magnetization. Also, by increasing the size of the network and the number of neurons in the hidden layer, it was possible to predict the critical temperature for the hexagonal Ising model. The closest value to the real thermodynamic value was obtained for the system size L = 80, corresponding to $T_c = 1.546 J/K_B$.

It is argued then, that the neural networks encode information of phase transitions by learning the order parameter, without knowledge of the Hamiltonian that represents the model.

References

[1.1] L. Wang, *Discovering Phase Transitions with Unsupervised Learning.* Beijing National Lab for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China (2016).

[1.2] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross*, Phys. Rev. B 89, 205118 (2014).*

[1.3] J. Snyder, M. Rupp, K. Hansen, K. Muller, and K. Burke. Physical review letters, 108, 253002 (2012).

[1.4] M. Rupp, A. Tkatchenko, K. Müller, and O. Anatole von Lilienfeld, Phys. Rev. Lett. 108, 058301 (2012).

[1.5] L. Onsager. *A two- dimensional model with an order-disorder transition*, Physical Review 65 117 (1944).

[1.6] B. M. McCoy and J.-Marie Maillard The Importance of the Ising Model Progress of Theoretical Physics, Vol. 127, No. 5, May 2012.

[1.7] S. Wetszel, Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders, Phys. Rev. E, 96, 022140 (2017).

[1.8] C. Wang and H. Zhai Machine learning of frustrated classical spin models. I. Principal component analysis, physical review B 96, 144432 (2017).

[1.9] D. Sanders Introducción a las transiciones de fase y a su simulación, Universidad Nacional Autónoma de México.

[2.1] Stephen G Brush, History of the Lenz-Ising model, Review of modern physics, vol. 39,4 19967.

[2.2] L. Onsager. A two- dimensional model with an order-disorder transition, Physical Review 65 117 (1944).

[2.3] Stephen L. Eltinge, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139-4307, (2015).

[2.4] A. Rebelo, Unsupervised learning of physical models: Uses and limitations of Principal Components Analysis, master thesis, Institute of Theorical Physics, (2017).

[2.5] Jean Claude Le Guillou and Jean Zinn-Justin. Accurate critical exponents from field theory. Journal de Physique, 50(12):1365–1370, 1989.

[2.6] L. Kadanoff, Phases of Matter and Phase Transitions; From Mean Field Theory to Critical Phenomena, The Perimeter Institute Waterloo, The James Franck Institute, 2009.

[2.7] MEJ Newman and GT Barkema. Monte Carlo Methods in Statistical Physics chapter 1-4. Oxford University Press: New York, USA, 1999.

[2.8] A. Rebelo, Unsupervised learning of physical models: Uses and limitations of Principal Components Analysis, master thesis, Institute of Theorical Physics, (2017).

[2.9] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 1953 J. Chem. Phys. 21, 1087.

[3.1] Stanford University (Producer). (2017). *Machine Learning*. Retrieved from https://www.coursera.org/learn/machine-learning/home/welcome.

[3.2] Lindsay I Smith, A tutorial on Principal Components Analysis, (2010).

[4.1] L. Wang, Discovering Phase Transitions with Unsupervised Learning, Phys. Rev. B, 94, 195105 (2016).

[4.2] J. Carrasquilla and R. Melko, Machine learning phases of matter, Nat. Phys. 13, 431-434 (2017).

[4.3] Diederik P. Kingma, Jimmy Ba, Adam: A Method for Stochastic Optimization, (2014).

[4.4] W. Hu, R. Singh and R. Scalettar, Discovering Phases, Phase Transitions and Crossovers through Unsupervised Machine Learning: A critical examination, Phys. Rev. E 95, 062122 (2017).

[4.5] L. Onsager. A two- dimensional model with an order-disorder transition, Physical Review 65 117 (1944).

[4.6] C. Wang and H. Zhai, Unsupervised Learning of Frustrated Classical Spin Models I: Principle Component Analysis, Phys. Rev. B, 96, 144432 (2017).

[4.7] A. Rebelo, Unsupervised learning of physical models: Uses and limitations of Principal Components Analysis, master thesis, Institute of Theorical Physics, (2017).

[4.8] D Kim and Dong-Hee Kim Smallest Neural network to learn the Ising Criticality, Phys. Rev. E, 98, 022138 (2018).