Critical Temperature from Unsupervised Deep Learning Autoencoders

based on Eur. Phys. J. B 93 (2020) 12, 226
and new results

Andreas Athenodorou

The Cyprus Institute & University of Pisa
https://orcid.org/0000-0003-4600-4245

Phase transitions in particle physics
Galileo Galilei Institute
Firenze

The Cyprus Institute RESEARCH·TECHNOLOGY·INNOVATION

NI4OS National Initiatives for Open Science in Europe

INFN Istituto Nazionale di Fisica Nucleare

Università di Pisa

Outline

Motivation

General goals

- Can we observe phase transitions in an unsupervise manner?

→ Using autoencoders

- Can this bring new information/tools into the game?

Physics goals

- Conventional MCMC algorithms.

→ Critical slowing down.

→ Difficulty in pinpointing critical temperature T_C.

- Observables with Finite Volume effects.

Algorithmic goals

- Understand domain of applicability of autoencoders.
- What are the limitations of autoencoders?

Motivation

General goals

- Can we observe phase transitions in an unsupervise manner?

→ Using autoencoders

- Can this bring new information/tools into the game?

Physics goals

- Conventional MCMC algorithms.

→ Critical slowing down.

→ Difficulty in pinpointing critical temperature T_C.

- Observables with Finite Volume effects.

Algorithmic goals

- Understand domain of applicability of autoencoders.
- What are the limitations of autoencoders?

Motivation

General goals

- Can we observe phase transitions in an unsupervise manner?

$\rightarrow$ Using autoencoders

- Can this bring new information/tools into the game?

Physics goals

- Conventional MCMC algorithms.

$\rightarrow$ Critical slowing down.

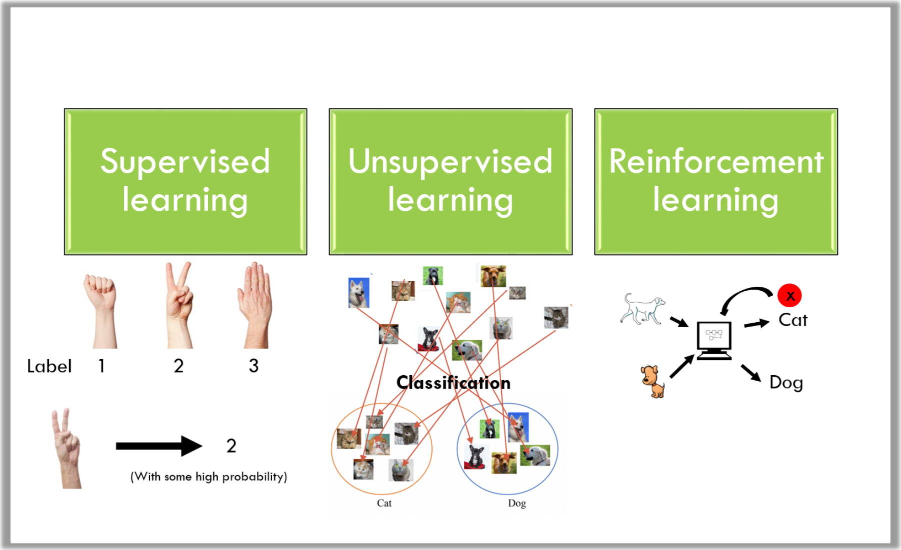
$\rightarrow$ Difficulty in pinpointing critical temperature T_C.

- Observables with Finite Volume effects.

Algorithmic goals

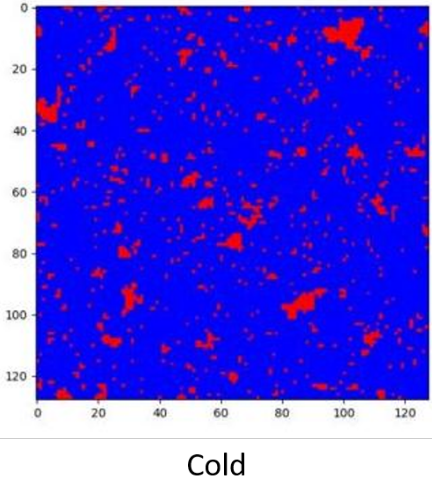
- Understand domain of applicability of autoencoders.
- What are the limitations of autoencoders?

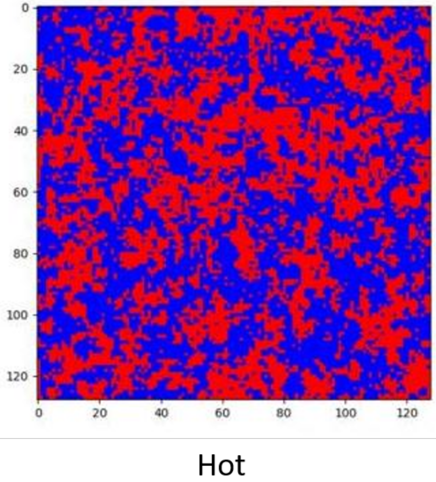
Machine Learning



Phase transitions as patterns of structure

Magnetic - Paramagnetic transition in 2D-Ising Model:





Phase transitions as patterns of structure

Can we identify the different phases using ML techniques?

- Supervised

  [Carrasquilla, Melko, Lucini, Wetzel, Lombardo ...]

- Unsupervised

  [Wetzel, Giataganas, Singh, ...]

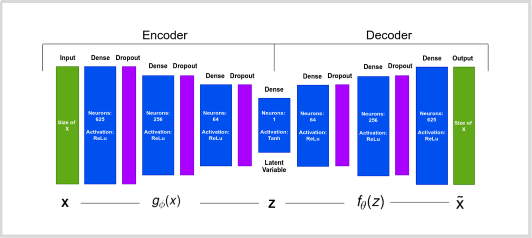
Can we extract thermodynamic quantities?

Can this define new observables?

Deep Learning Autoencoders

Objective: Learning features in a given dataset hierarchically.

- Autoencoders (AE): Dimensionality reduction.

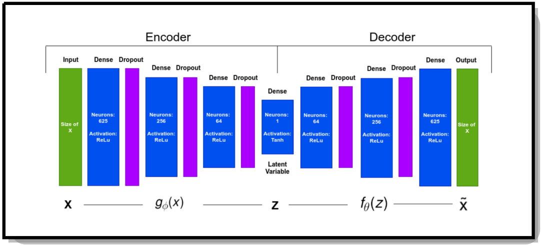


- An autoencoder encodes the input data ($\{X\}$) from the input layer into a latent variable ($\{z\}$).
- Then it uncompresses that latent variable into an approximation of the original data ($\{\tilde{X}\}$).
- Learns to ignore the noise and recognize significant characteristics of the input data.

Deep Learning Autoencoders

Objective: Learning features in a given dataset hierarchically.

- Autoencoders (AE): Dimensionality reduction.

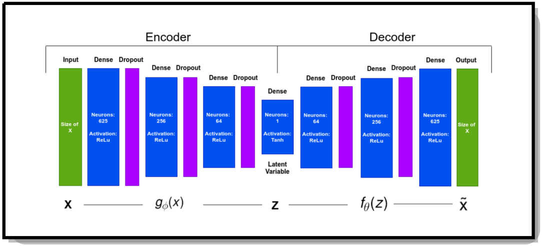


- An autoencoder encodes the input data ($\{X\}$) from the input layer into a latent variable ($\{z\}$).
- Then it uncompresses that latent variable into an approximation of the original data ($\{X\}$).
- Learns to ignore the noise and recognize significant characteristics of the input data.

Deep Learning Autoencoders

Objective: Learning features in a given dataset hierarchically.

- Autoencoders (AE): Dimensionality reduction.



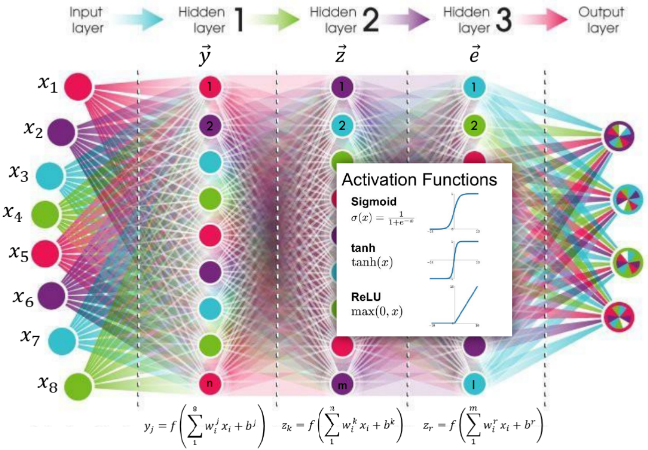
- An autoencoder encodes the input data ($\{X\}$) from the input layer into a latent variable ($\{z\}$).
- Then it uncompresses that latent variable into an approximation of the original data ($\{X\}$).
- Learns to ignore the noise and recognize significant characteristics of the input data.

Typical Neural Network

1. Input layer
2. Hidden layer
3. Hidden layer
4. Hidden layer
5. .
6. .
7. .
8. .
9. .

....

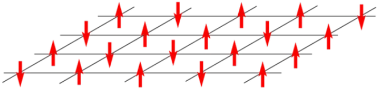
10. Output layer



Credits: neuralnetworksanddeeplearning.com 2016

Ferromagnetic Ising Model

- $1D$ Not so interesting: No phase transition (never magnetised)
- $2D$ more interesting: There is a phase transition
- Simplest Description of Ferromagnetism



- Hamiltonian:

$$H = -J \sum_{i,j=nn(i)}^{N} s_i s_j - \mu h \sum_{i=1}^{N} s_i$$

Nearest neighbors

Observables:

- Magnetization is the order parameter:

$$m = \frac{1}{N} \sum_{i=1}^{N} |s_i|$$

The 2D Ising model has a second order phase transition (magnetization is continuous)

- Magnetic susceptibility

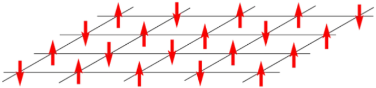
$$\chi = \frac{N}{T} (\langle m^2 \rangle - \langle m \rangle^2)$$

- Heat Capacity

$$C = \frac{\partial \langle E \rangle}{\partial T}$$

Ferromagnetic Ising Model

- $1D$ Not so interesting: No phase transition (never magnetised)
- $2D$ more interesting: There is a phase transition
- Simplest Description of Ferromagnetism



- Hamiltonian:

$$H = -J \sum_{i,j=nn(i)}^{N} s_i s_j - \mu h \sum_{i=1}^{N} s_i$$

Nearest neighbors

Observables (near criticality $\sim T_c$):

- Magnetization is the order parameter:

$$m = \frac{1}{N} \sum_{i=1}^{N} |s_i| \qquad m(T) \sim |T - T_c|^b$$

The 2D Ising model has a second order phase transition (magnetization is continuous)

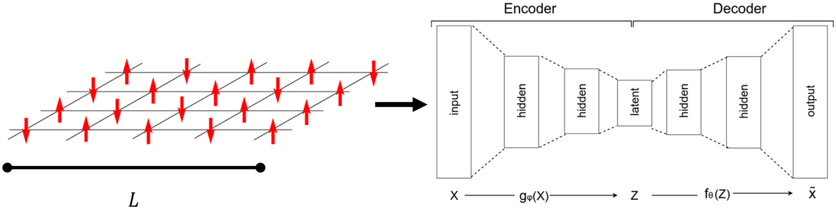
- Magnetic susceptibility

$$\chi = \frac{N}{T} (\langle m^2 \rangle - \langle m \rangle^2) \qquad \chi(T) \sim |T - T_c|^{-\gamma}$$

- Heat Capacity

$$C = \frac{\partial \langle E \rangle}{\partial T} \qquad \chi(T) \sim |T - T_c|^{-\alpha}$$

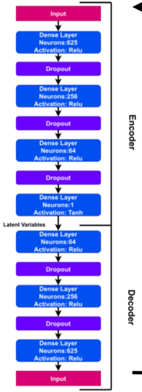
Procedure

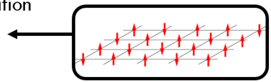


- The autoencoder aims to define a representation (encoding) for our assemblage of data, by performing dimensionality reduction:
  - It encodes the input data ($\{X\}$) from the input layer into a latent variable ($\{z\}$)
  - Then it uncompresses that latent variable into an approximation of the original data ({X})
- It consists of two components, the encoder function g_φ and a decoder function f_θ and the reconstructed input is $X = f_\theta(g_\varphi(x))$
- The autoencoder learns the parameters φ and θ together
- $X = f_\theta(g_\varphi(x))$ can approximate an identity function.
- Criterion: Minimize the Mean Square Error (MSE): $MSE(\theta, \phi) = \frac{1}{n_{\text{data}}} \sum_{i=1}^{n_{\text{data}}} (X_i - f_\theta(g_\phi(X_i)))^2 .$

Procedure

- Eight layers
- Fully connected (Dense)
- **Single** latent variable
- Through experimentation



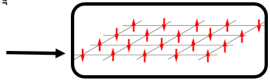


$$\text{relu} : y = max(0, x) = \begin{cases} x, \text{ if } x > 0 \\ 0 \quad \text{if } x \leq 0 \end{cases}$$

$$\tanh : y = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

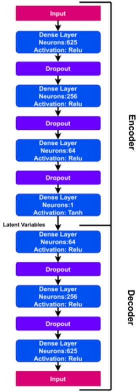
Dropout, reduces overfitting

G. E. Hinton et al, arXiv:1207.0580.



Procedure

- Eight layers
- Fully connected (Dense)
- **Single** latent variable
- Through experimentation



Implementation:

- Using Keras (F. Chollet et al., "Keras." https://keras.io, 2015)
- Using Tensorflow (M. Abadi et al OSDI, vol. 16, pp. 265–283, 2016.)

- 66.66.% of data used for training
- 33.33.% of data used for validating
- Training is performed for 2000 iterations

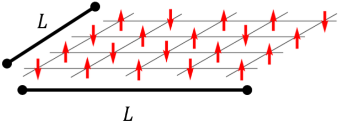
- Example:

```
input_img = Input(shape=(original_dim,))
encoded = Dense(2048, activation='tanh')(input_img)
encoded = Dropout(0.4)(encoded)
encoded = Dense(512, activation='tanh')(encoded)
encoded = Dropout(0.4)(encoded)
encoded = Dense(64, activation='tanh')(encoded)
encoded = Dropout(0.4)(encoded)
encoded = Dense(latent_dim, activation='linear')(encoded)

decoded = Dense(64, activation='tanh')(encoded)
decoded = Dropout(0.4)(decoded)
decoded = Dense(512, activation='tanh')(decoded)
decoded = Dropout(0.4)(decoded)
decoded = Dense(2048, activation='tanh')(decoded)
decoded = Dropout(0.4)(decoded)
decoded = Dense(original_dim, activation='linear')(decoded)
```

Procedure

- Each configuration is re-expressed in the form of a vector:



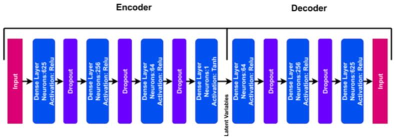
→

L^2

(-1, 1, -1,1,1, -1,1,-1,1,-1,1,1,-1,1,-1,1,1,1,1,1,-1)

↓

Input into the encoder



↓

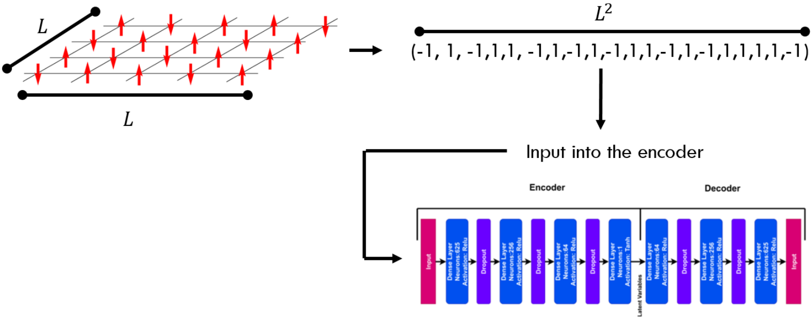
$z_{i_{\mathrm{conf}}}$

Latent Variable

- In other words, each configuration is assigned a number, the latent variable, which includes all the physically necessary information so that the decoder re-creates the actual configuration

Procedure

- Each configuration is re-expressed in the form of a vector:



- In other words, each configuration is assigned a number, **the latent variable,** which includes all the physically necessary information so that the decoder re-creates the actual configuration

- The autoencoder receives information for only one lattice volume, and thus, it "knows" nothing about configurations produced for other volume sizes

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.

- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.

- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:

  $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.
- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.
- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:
  - $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.

- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.

- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:

  $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.

- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.

- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:

  $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.
- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.
- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:

  $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

Procedure

Attempt to identify signals of the phase structure of the 2D-Ising model

Question: How the latent dimension behaves as a function of the temperature T for each configuration.

- We produce in total 40000 configurations.
- 200 configurations for each single temperature T.
- For the temperature range $T = 1 - 4.5$.

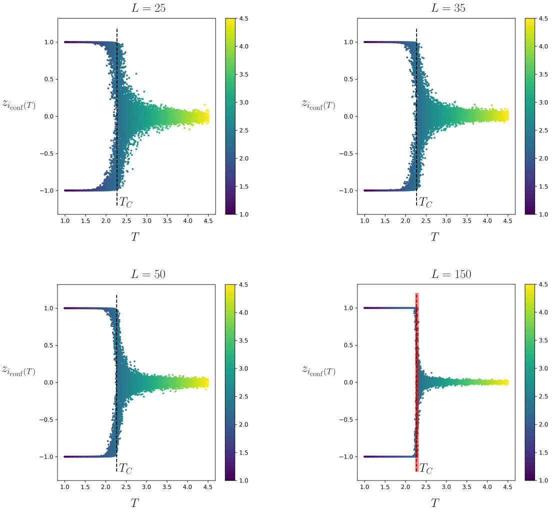
- We make sure that we cover the whole range of temperatures between the two extreme cases of the ising behaviour:
  - The nearly "frozen" at $T = 1$.
  - The disordered at $T = 4.5$.
  - We assume that we have no prior knowledge on what is happening in-between.

- Wait! Isn't there a degree of supervision?
  - We could have chosen different temperature ranges:

  $\rightarrow$ For instance $T = 0.01 - 1000$ (Computationally more expensive).

The latent variable per configuration

We plot the latent variable for each different configuration, as a function of the temperature T for different L.



$40,000$ configs, $2/3$ training, $1/3$ validation.

The latent variable per configuration

We plot the latent variable for each different configuration, as a function of the temperature T for different L.

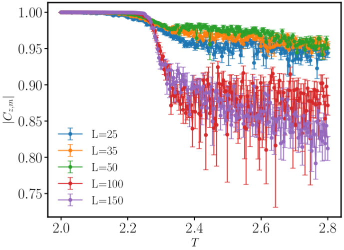
For low temperatures we obtain two plateaus, one located at $z = 1$ and one at $z = -1$.

- Corresponds to two distinct states that are not connected through any kind of transformation.
- This reflects the spontaneously broken $Z_2 \equiv \{-1, 1\}$ global symmetry group.

One can interpret these two plateaus as the two cases where all spins are up or down.

- Is this true?
- We test the correlation coefficient between the latent variable to magnetization.

$$C_{z,m} = \frac{\langle (z - \bar{z})(m - \bar{m}) \rangle}{\sqrt{\langle (z - \bar{z})^2 \rangle \langle (m - \bar{m})^2 \rangle}} .$$



The latent variable per configuration

We plot the latent variable for each different configuration, as a function of the temperature T for different L.

For low temperatures we obtain two plateaus, one located at $z = 1$ and one at $z = -1$.

- Corresponds to two distinct states that are not connected through any kind of transformation.
- This reflects the spontaneously broken $Z_2 \equiv -1, 1$ global symmetry group.

One can interpret these two plateaus as the two cases where all spins are up or down.

- Is this true?
- We test the correlation coefficient between the latent variable to magnetization.
- The latent variable -1 and 1 corresponds to the two orientations of the spins.
- Finally, the two plateaus become more distinct as the lattice size increases.

At some temperature range ΔT_{Trans} the aforementioned behaviour collapses to one state, which is located around $z = 0$. This reflects the restoration of Z_2 symmetry.

- In other words, it corresponds to the case where all the spins are disoriented.

Critical point: as the lattice size increases the width of this transition decreases and this step becomes steeper and steeper. At $L \to \infty$, $T_C(L)$ is localised right on the critical temperature T_C.

Different Temperature Windows

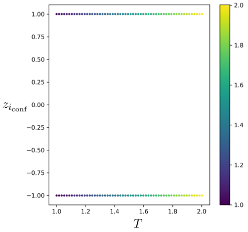
What happens within different "temperature windows"

What if we use a temperature window within the range $T = 1 - 2$ and apply the autoencoder

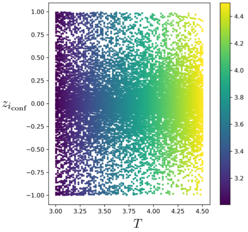
- Two ordered states are visible without the presence of a critical point.
- Since there is no visible signal for a phase transition within this range of T it is reasonable to use another temperature window.

What if we choose $T = 3 - 4.5$?

- No particular pattern is observed



$T = 1 - 2$

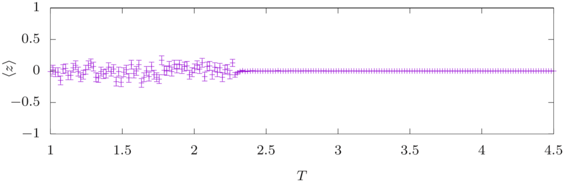


$T = 3 - 4.5$

Next step would be to go to $T = 2 - 3$

Can we define macroscopic quantities?

First: Average latent variable $\langle z \rangle$ as a function of the Temperature T.



For low temperatures the latent dimension is, in a good approximation, equally distributed between the values of -1 and 1.

Safe to consider absolute value of the latent variable.

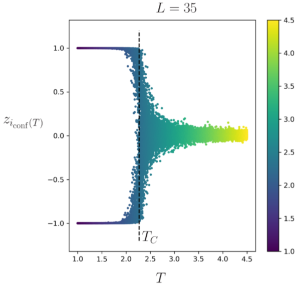
Can we define macroscopic quantities?

Since the latent dimension per confguration is symmetric with respect to the axis, it would be reasonable to define the average absolute latent variable

$$\tilde{z} = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} |z_{i_{\text{conf}}}| \ .$$

The latent variable resembles the behaviour of the magnetization per spin configuration as a function of the temperature:

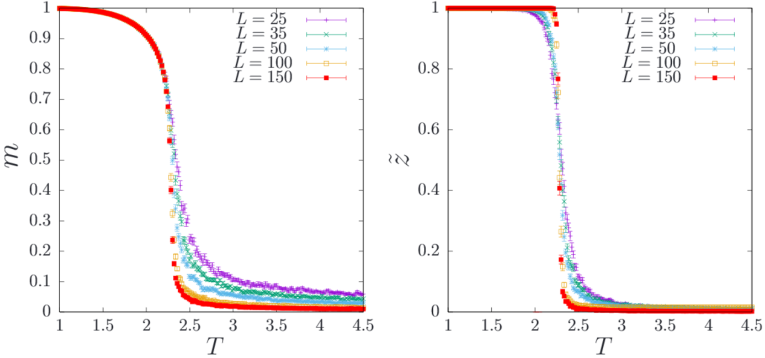
$$m = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} |s_{i_{\text{conf}}}| \ .$$



Latent Susceptibility & Critical Temperature

We print the average absolute latent variable as a function of the temperature

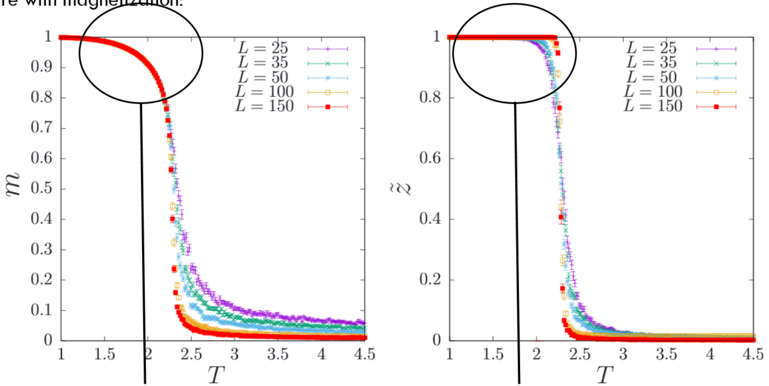
Compare with magnetization:



Latent Susceptibility & Critical Temperature

We print the average absolute latent variable as a function of the temperature

Compare with magnetization:

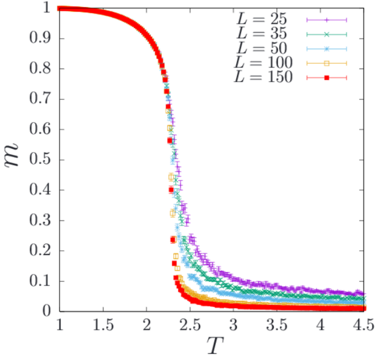


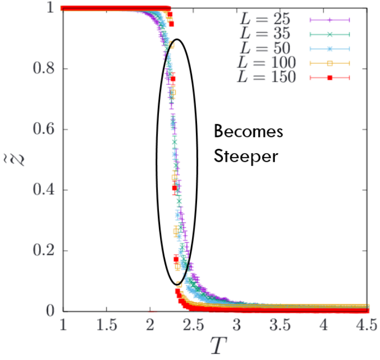
Absolute average latent variable can be used as an order parameter to identify the critical temperature, but cannot capture the right order of the phase transition

Latent Susceptibility & Critical Temperature

We print the average absolute latent variable as a function of the temperature

Compare with magnetization:





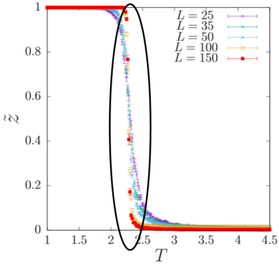
$T_c(L)$ extracted from the autoencoder data will suffer less from finite size scaling effects

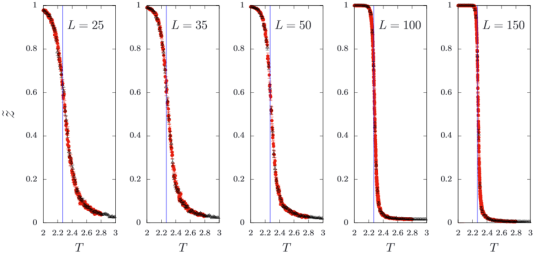
Latent Susceptibility & Critical Temperature

We define the latent susceptibility as $\chi_{\tilde{z}} = \frac{L^2}{T}\left(\langle \tilde{z}^2\rangle - \langle \tilde{z}\rangle^2\right)$.

Like the magnetic susceptibility $\chi = \frac{L^2}{T}\left(\langle m^2\rangle - \langle m\rangle^2\right)$.

Their peaks determine $T_c(L)$





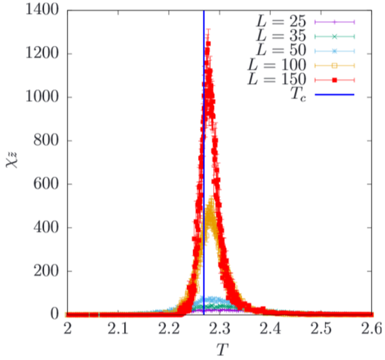
More points are required

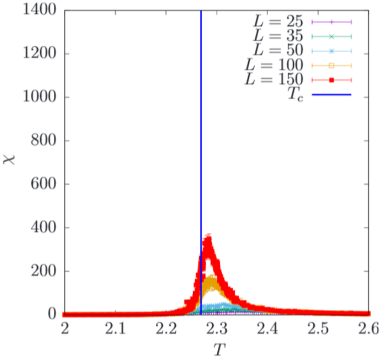
(Using the extracted weights)

No overfitting!!!!

Latent Susceptibility & Critical Temperature

We extract the latent susceptibility as well as the magnetic susceptibility:



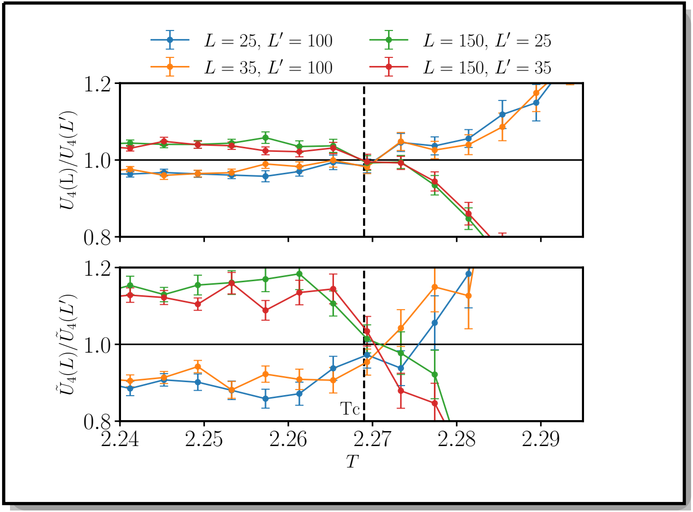


From the peaks we extract $T_c(L)$

Does it extrapolate to the right value $T_c = 2.269184$?

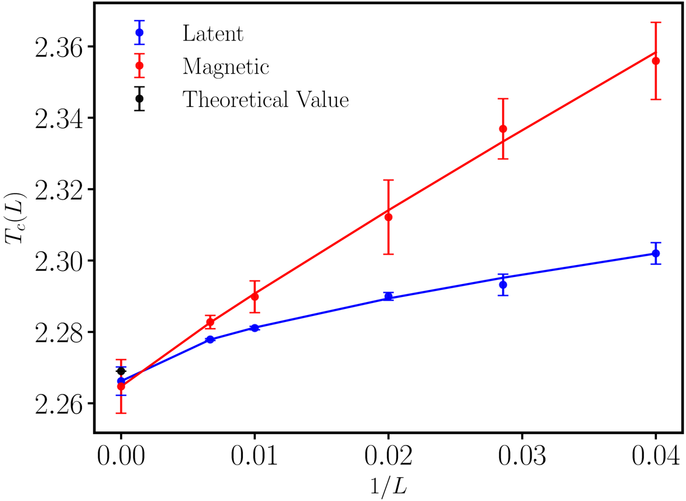
Extracting the T_C using Binder cumulants

$$U_4 = 1 - \frac{\langle m^4 \rangle}{3\langle m^2 \rangle^2} \qquad \tilde{U}_4 = 1 - \frac{\langle \tilde{z}^4 \rangle}{3\langle \tilde{z}^2 \rangle^2}$$



Extracting the T_C in the Thermodynamic Limit

- Noisy Binder Cumulant ratios, first indication that issues in Finite Size Scaling.
- $T_C(L)$ from peaks of latent and magnetic susceptibility as a function of $1/L$.



Takebacks

$$T_c(L) - T_c(L=\infty) \propto L^{-1/\nu}$$

Susceptibility	$T_c(L=\infty)$	ν	$\chi^2/$ dof
Magnetic	$2.265(8)$	$1.08(20)$	0.15
Latent	$2.266(4)$	$1.60(14)$	0.41

- Critical temperature can be extracted to adequate accuracy.
- Observed $\mathcal{Z}_2$ symmetry broken.
- Configurations from latent variable are from a different universaity class, but share the same $T_C(\infty)$.
- Latent variable suffers from small finite volume effects, can help in constructing observables with small FV effects.

The anti-ferromagnetic Ising model

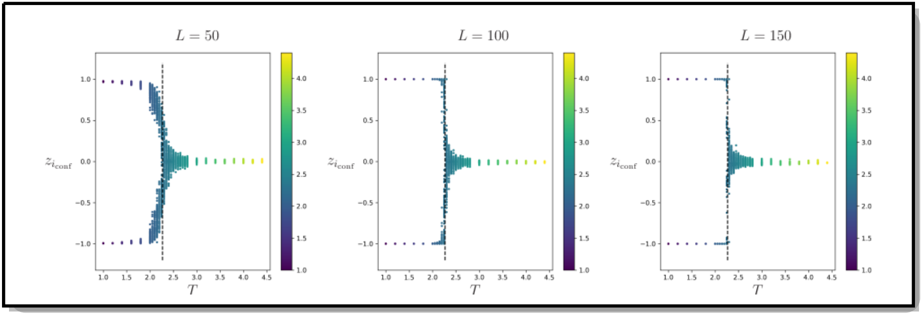
Can we use the same methodology on the 2D anti-ferromagnetic Ising model?

- We generate 6000 configurations, namely 200 configurations for every single temperature
- The configurations are for 30 different values of temperature within the range $T = 1 - 4.5$.

Expectation: Since anti-ferromagnetic is connected with ferromagnetic via a bijective map:

$\rightarrow$ The autoencoder should be able to "notice" the phase transition

Result: Latent variable for each configuration:



The anti-ferromagnetic Ising model

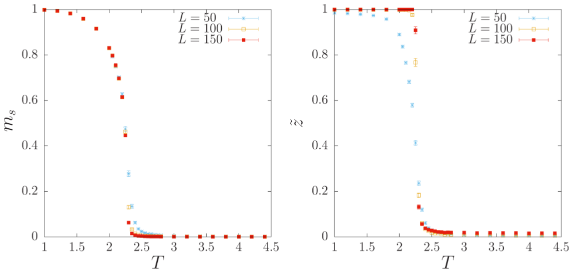
Can we use the same methodology on the 2D anti-ferromagnetic Ising model?

- We generate 6000 configurations, namely 200 configurations for every single temperature
- The configurations are for 30 different values of temperature within the range $T = 1 - 4.5$.

Expectation: Since anti-ferromagnetic is connected with ferromagnetic via a bijective map:

$\rightarrow$ The autoencoder should be able to "notice" the phase transition

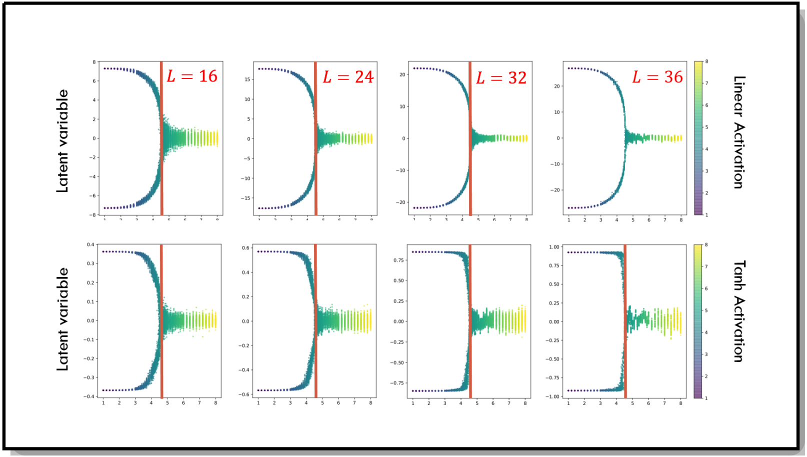
Result: Average absolute latent variable:



3D Ising Model

Critical temperature $T_C = 4.511$, Second order.

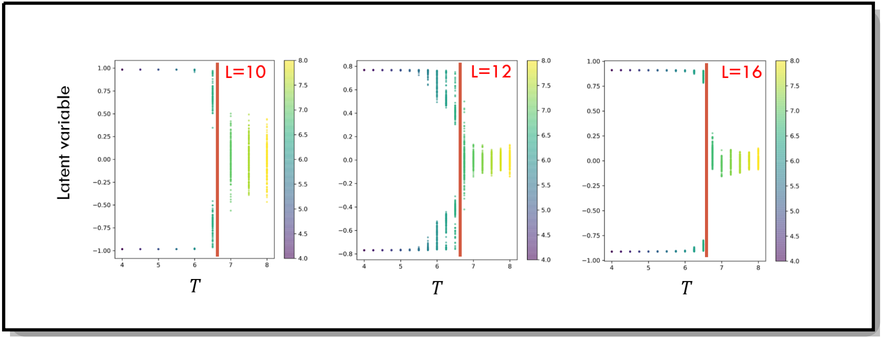
[Talapov & Blöte 1996]



4D Ising Model

Critical temperature $T_C = 6.65$.

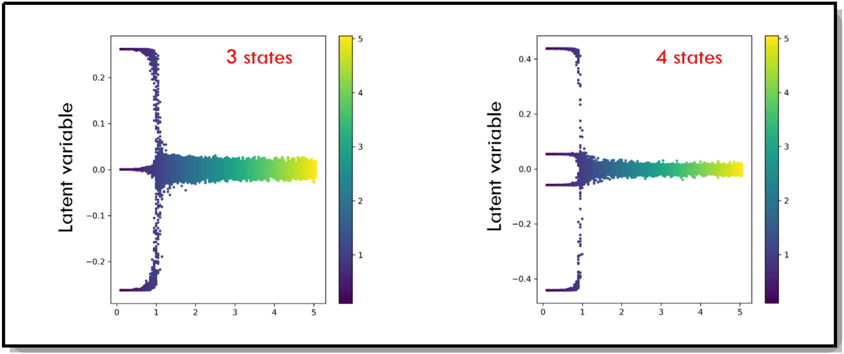
[Lundow & Markström 2012]



Potts Model

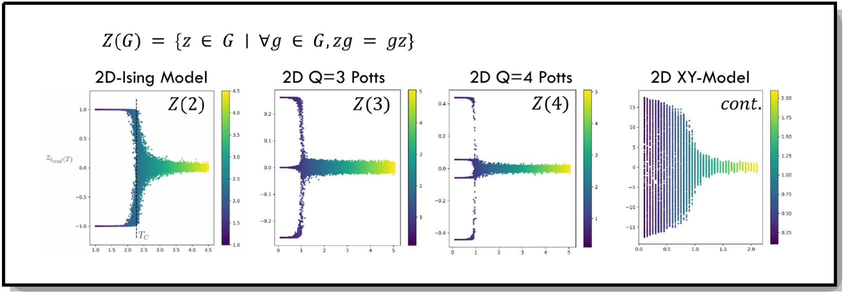
$q \leq 4$ second order, $q > 4$ first order.

[Wu 1982]



Summary

- We define new observables: The latent variable and latent susceptibility
- We can determine the critical temperature T_C to an adequate accuracy.
- We can extract T_C as a function of L with small finite scaling effects.



- Autoencoders detect broken center symmetry of the underlying group.
- Significant effects of the choice of activation functions on the order of the phase transition.

Outlook

- Need to test on theories whose order parameters are not a moment of the field variable.
- What information more latent variables provide?
- Looking forward to gauge theories:
  - Z_2
  - Compact $U(1)$ (infinite order phase transition.)
  - $SU(N)$ (second order for small N, first order as we increase N.)
- Supervised/unsupervised investigation of Topological field objects in statistical systems and quantum field theories (Instantons, Nielsen-Olesen vortices)

Thanks to all my collaborators: C. Alexandrou, C. Chrysostomou. S. Paul, A. Apseros, C. Havadjia, S. Shiakas and D. Vadacchino.

Thanks for your attention!!!