Today

Lecture 1
Using Machine Learning to Explore Neural Data

Lecture 2
Magnets, Machines, Brains

Friday

Lecture 3
Deep Learning
Daniel Shechtman
2011 Nobel Laureate
in Chemistry

George Smoot
2006 Nobel Laureate
in Physics
Outlook

Magnets & Ising model

Computational problems

Artificial neural networks

Retina
Magnetism
Magnetism
Magneto?

Wolf Singer
(former director Max-Planck Institute)
jueves, 15 de octubre de 15
jueves, 15 de octubre de 15
Ferromagnets

Some materials are permanent magnets: Iron, Cobalt, Nickel, alleys,...
Ferromagnets

Some materials are **permanent** magnets: Iron, Cobalt, Nickel, alleys,...

**Electromagnetism**: physical interaction between electrically charged particles
Ferromagnets

Some materials are **permanent** magnets: Iron, Cobalt, Nickel, alleys,...

**Electromagnetism:** physical interaction between electrically charged particles

Electric field: stationary charges
Ferromagnets

Some materials are **permanent** magnets: Iron, Cobalt, Nickel, alleys,...

**Electromagnetism:** physical interaction between electrically charged particles

Electric field: stationary charges

Magnetic field: moving charges (currents)
Electrons in a material have an orbital magnetic moment and an intrinsic spin.
**Ferromagnets**

**Electrons** in a material have an orbital magnetic moment and an intrinsic **spin**

---

**Classic Physics view!**
Ferromagnets

In Ferromagnets spins tend to be parallel to each other & to an applied magnetic field
Ferromagnets

In Ferromagnets spins tend to be parallel to each other & to an applied magnetic field

Can the local interactions align enough spins to produce a global field?
Ferromagnets

In Ferromagnets spins tend to be parallel to each other & to an applied magnetic field.

Can the local interactions align enough spins to produce a global field?

Spin alignment Vs. Temperature
**Model**

**Ingredient 1:** atoms are arranged in a lattice (crystal solid)
Model

**Ingredient 1:** atoms are arranged in a lattice (crystal solid)

\[ \Lambda^d \]

\[ \sigma_i = \pm 1 \quad \sigma = (\sigma_1 \ldots \sigma_N) \]
Model

**Ingredient 1:** atoms are arranged in a lattice (crystal solid)

**Ingredient 2:** energy as a function of spins alignment (magnetic)

\[ \sigma_i = \pm 1 \quad \sigma = (\sigma_1 \ldots \sigma_N) \]
**Model**

**Ingredient 1:** atoms are arranged in a lattice (crystal solid)

**Ingredient 2:** energy as a function of spins alignment (magnetic)

\[
\Lambda^d
\]

\[
\sigma_i = \pm 1 \quad \sigma = (\sigma_1 \ldots \sigma_N)
\]

\[
H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i
\]
Model

**Ingredient 1:** atoms are arranged in a lattice (crystal solid)

**Ingredient 2:** energy as a function of spins alignment (magnetic)

**Ingredient 3:** Gibbs probability measure over states (temperature)

\[
\sigma_i = \pm 1 \quad \quad \sigma = (\sigma_1 \ldots \sigma_N)
\]

\[
H(\sigma) = -J \sum_{\{i,j\}} \sigma_i \sigma_j - h \sum_i \sigma_i
\]
**Ingredient 1**: Atoms are arranged in a lattice (crystal solid).

**Ingredient 2**: Energy as a function of spins alignment (magnetic).

**Ingredient 3**: Gibbs probability measure over states (temperature).

\[ \sigma_i = \pm 1 \quad \sigma = (\sigma_1 \ldots \sigma_N) \]

\[ H(\sigma) = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \quad \beta = \frac{1}{\kappa_B T} \]
A negative thesis?

**I-D** (Ising 1924): temperature always wins the battle and concluded that no phase transition existed
A negative thesis?

1D (Ising 1924): temperature always wins the battle and concluded that no phase transition existed

2D (Peierls 1936; Onsager 1944): critical $T$ below which there is net magnetization. Phase transition!
A negative thesis?

**1D** (Ising 1924): temperature always wins the battle and concluded that no phase transition existed

**2D** (Peierls 1936; Onsager 1944): critical $T$ below which there is net magnetization. Phase transition!

\[ M = \langle \sum_{i \in \Lambda} \sigma_i \rangle \]
A negative thesis?

1D (Ising 1924): temperature always wins the battle and concluded that no phase transition existed

2D (Peierls 1936; Onsager 1944): critical $T$ below which there is net magnetization. Phase transition!

$$M = \left\langle \sum_{i \in \Lambda} \sigma_i \right\rangle$$
How local interactions can give rise to global phenomena?

**Collective phenomena:** repetitive interactions among many individuals produce patterns on a scale larger than themselves.
How local interactions can give rise to global phenomena?

**Collective phenomena:** repetitive interactions among many individuals produce patterns on a scale larger than themselves
How local interactions can give rise to global phenomena?

**Collective phenomena:** repetitive interactions among many individuals produce patterns on a scale larger than themselves
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

N is typically a large number: one needs to sample from a very high-dimensional probability distribution!
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\( \sigma = (\sigma_1 \ldots \sigma_N) \)

\( N \) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\( N \) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\( N \) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

**MCMC methods:** Metropolis, heat-bath algorithm (Glauber),...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\(N\) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Criticality

At critical temperature exists a regime **between order** and **disorder** with very special statistical properties.
Criticality

Divergence correlation length $\langle \sigma_i \sigma_j \rangle \propto \exp(-|i - j|/\zeta)$
Divergence correlation length \( \langle \sigma_i \sigma_j \rangle \propto \exp(-|i - j|/\zeta) \)

Power-law divergences \( P(x) \propto x^{-\alpha} \)
Criticality

Divergence correlation length \( \langle \sigma_i \sigma_j \rangle \propto \exp(-|i-j|/\zeta) \)

Power-law divergences \( P(x) \propto x^{-\alpha} \)

Maximal sensitivity \( C = \frac{\sigma_E^2}{kT^2} \)
Criticality

Divergence correlation length
\[ \langle \sigma_i \sigma_j \rangle \propto \exp\left(-\frac{|i - j|}{\zeta}\right) \]

Power-law divergences
\[ P(x) \propto x^{-\alpha} \]

Maximal sensitivity
\[ C = \frac{\sigma_E^2}{kT^2} \]

Fractal dimensions
Criticality

2D systems at criticality may exhibit **conformal invariance** that can be exploited to obtain rigorous results about interfaces and random curves (self-avoiding random walks, percolation, Ising, Gaussian-free fields...)

Wendelin Werner
(Fields medal 2006)

Stanislav Smirnov
(Fields medal 2010)
I slide course on Information theory
## Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static</strong></td>
<td><strong>Shannon entropy</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Dynamic</strong></td>
<td><strong>Kolmogorov entropy</strong></td>
<td><strong>Transfer entropy</strong></td>
</tr>
<tr>
<td>Static</td>
<td>Bivariate</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td><em>Shannon entropy</em></td>
<td></td>
</tr>
<tr>
<td>Dynamic</td>
<td><em>Kolmogorov entropy</em></td>
<td></td>
</tr>
</tbody>
</table>
# Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Shannon entropy</td>
<td>$I_{i_1 \ldots i_n i_{n+1} i_M}$</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Kolmogorov entropy</td>
<td>$J_{j_1 \ldots j_n j_{n+1} j_M}$</td>
</tr>
</tbody>
</table>
## Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Shannon entropy</td>
<td>$p(i,j) = p(i)p(j)$</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Kolmogorov entropy</td>
<td>$i_1 \ldots i_n \ i_{n+1} \ i_M$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$j_1 \ldots j_n \ j_{n+1} \ j_M$</td>
</tr>
</tbody>
</table>
# Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Shannon entropy</td>
<td>Mutual information</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Kolmogorov entropy</td>
<td></td>
</tr>
</tbody>
</table>

\[
p(i,j) = p(i)p(j)
\]
## Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Shannon entropy</td>
<td>Mutual information</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Kolmogorov entropy</td>
<td></td>
</tr>
</tbody>
</table>

\[
p(i,j) = p(i)p(j)
\]

\[
p(i_{n+1}|i_n,j_n) = p(i_{n+1}|i_n)
\]
### Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static</strong></td>
<td>Shannon entropy</td>
<td>Mutual information</td>
</tr>
<tr>
<td><strong>Dynamic</strong></td>
<td>Kolmogorov entropy</td>
<td>Transfer entropy</td>
</tr>
</tbody>
</table>

\[
p(i, j) = p(i)p(j)
\]

\[
p(i_{n+1} | i_n, j_n) = p(i_{n+1} | i_n)
\]
### Information theory

<table>
<thead>
<tr>
<th>Static</th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shannon entropy</strong></td>
<td></td>
<td><strong>Mutual information</strong></td>
</tr>
<tr>
<td><strong>Kolmogorov entropy</strong></td>
<td></td>
<td><strong>Transfer entropy</strong></td>
</tr>
</tbody>
</table>

\[
p(i, j) = p(i)p(j)
\]

\[
p(i_{n+1} | i_n, j_n) = p(i_{n+1} | i_n)
\]

---

Measuring Information Transfer

Thomas Schreiber

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, 01187 Dresden, Germany

(Received 19 January 2000)

An information theoretic measure is derived that quantifies the statistical coherence between systems evolving in time. The standard time delayed mutual information fails to distinguish information that is actually exchanged from shared information due to common history and input signals. In our new approach, these influences are excluded by appropriate conditioning of transition probabilities. The resulting transfer entropy is able to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems.

\[
T_{j \rightarrow I} = \sum p(i_{n+1}, i_n^{(k)}, j_n^{(l)}) \log \frac{p(i_{n+1} | i_n^{(k)}, j_n^{(l)})}{p(i_{n+1} | i_n^{(k)})}.
\]
# Information theory

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Shannon entropy</td>
<td>Mutual information</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Kolmogorov entropy</td>
<td>Transfer entropy</td>
</tr>
</tbody>
</table>

\[ p_{ij} = p(i)p(j) \]

\[ p_{ij+1|i_n,j_n} = p(i_{n+1}|i_n) \]

**Transfer entropy** measures information exchange rather than information sharing.
**Criticality**

**Mutual information** is maximal at criticality

Information flow in a kinetic Ising model peaks in the disordered phase

Criticality

**Transfer entropy** peaks at the disordered phase!

Information flow in a kinetic Ising model peaks in the disordered phase

Outlook

Magnets & Ising model

Computational problems

Artificial neural networks

Retina
Computational complexity

**Computability theory:** is there an algorithm that solves the problem?

**Computational complexity theory:** how much resources needs the algorithm?
Computational complexity

**Intuition:** classify the difficulty of problems by the amount of computational resources (time, memory,...) needed to solve them

How powerful needs to be a machine to solve a problem in a given time?

**Class P:** problems solved in polynomial time by a deterministic Turing machine

**Class NP:** problems solved in polynomial time by a non-deterministic Turing machine
Computational complexity

**Algorithms:** anything a Turing machine can do

**Local search algorithms:** algorithms that start from a candidate solution, look around at proximal candidates, make a decision about where to move on

E.g. gradient descent

...going down a mountain with a bike in a pitch black night
Computational problems

**Decision problems:** a family of instances (input) together with a solution for each instance (yes/no)

E.g. given a number $n$, is $n$ a prime number?
Computational problems

**Ex. 1** Number partition (NP-complete): given a set of natural numbers $S = \{n_1, \ldots, n_N\}$ can I find a partition in 2 sets with equal sum?
Computational problems

**Ex. 1** Number partition (NP-complete): given a set of natural numbers $S = \{n_1, \ldots, n_N\}$ can I find a partition in 2 sets with equal sum?

e.g. $\{3,1,1,2,2,1\} \rightarrow \{3,2\} \ {1,1,2,1\}$
Ex. 1 Number partition (NP-complete): given a set of natural numbers $S = \{n_1, \ldots, n_N\}$ can I find a partition in 2 sets with equal sum?

e.g. $\{3,1,1,2,2,1\} \rightarrow \{3,2\} \{1,1,2,1\}$

$$H = A \left( \sum_{i=1}^{N} n_i s_i \right)^2$$

$s_i = \pm 1$

$A > 0$
Computational problems

**Ex. 1** Number partition (NP-complete): given a set of natural numbers \( S = \{n_1, \ldots, n_N\} \) can I find a partition in 2 sets with equal sum?

\[
\text{e.g. } \{3,1,1,2,2,1\} \rightarrow \{3,2\}\ {\{1,1,2,1\}}
\]

\[
H = A \left( \sum_{i=1}^{N} n_i s_i \right)^2
\]

\[
s_i = \pm 1
\]

\[
A > 0
\]

\[
H = 0?
\]
Computational problems

**Ex. 2** Graph coloring (NP-complete): given a graph $G = (V, E)$ and a set of $n$ colors, is it possible to color each vertex such that no edge connects two vertices of same color?
**Computational problems**

**Ex. 2** Graph coloring (NP-complete): given a graph $G = (V, E)$ and a set of $n$ colors, is it possible to color each vertex such that no edge connects two vertices of same color?

$$H = A \sum_v \left( 1 - \sum_{i=1}^{n} x_{v,i} \right)^2 + A \sum_{(uv) \in E} \sum_{i=1}^{n} x_{u,i} x_{v,i}.$$

[Graph coloring diagram]
Computational problems

**Ex. 2** Graph coloring (NP-complete): given a graph $G = (V, E)$ and a set of $n$ colors, is it possible to color each vertex such that no edge connects two vertices of same color?

$$H = A \sum_v \left( 1 - \sum_{i=1}^n x_{v,i} \right)^2 + A \sum_{(uv) \in E} \sum_{i=1}^n x_{u,i} x_{v,i}.$$ 

$$H = 0?$$
Computational problems

**Ex. 3** Boolean satisfiability (3-SAT). Is there an assignation of variables to make a logic formula true?

\[(\bar{x} \lor y \lor z) \land (x \lor \bar{y} \lor z) \land (x \lor y \lor \bar{z}) \land (x \lor \bar{y} \lor \bar{z}) \land (\bar{x} \lor y \lor z) \land (\bar{x} \lor \bar{y} \lor \bar{z})\]

\# variables = 3
\# clauses = 6
Computational problems

Many **NP-complete** and NP-hard problems (including all 21 Karp’s NP-complete) can be formulated as quadratic (Ising) energy functions

\[ H(\sigma) = -\sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i \]

**g.s.** \( H \leq 0 \)?

**Ising formulations of many NP problems**
Computational problems

Many **NP-complete** and NP-hard problems (including all 21 Karp’s NP-complete) can be formulated as quadratic **(Ising)** energy functions

\[ H(\sigma) = - \sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i \quad \text{g.s. } H \leq 0? \]

**Ising formulations of many NP problems**

Problems of scientific interest can be **encoded** and approx. solved in **experimental devices** using Ising energy functions
Simulating Ising models

\[
H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i
\]

\[
P(\sigma) \propto e^{-\beta H(\sigma)}
\]

\[
\sigma = (\sigma_1 \ldots \sigma_N)
\]
Simulating Ising models

$H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i$

$P(\sigma) \propto e^{-\beta H(\sigma)}$

$\sigma = (\sigma_1 \ldots \sigma_N)$

$N$ is typically a large number: one needs to sample from a very high-dimensional probability distribution!
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\(N\) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\( N \) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\(N\) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber), ...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Simulating Ising models

\[ H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \]

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

\[ \sigma = (\sigma_1 \ldots \sigma_N) \]

\( N \) is typically a large number: one needs to sample from a very high-dimensional probability distribution!

MCMC methods: Metropolis, heat-bath algorithm (Glauber),...

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Computational problems

Endowing with **dynamics** the Ising energy functions associated with problems amounts to run a **local search** algorithm.

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]
Computational problems

Endowing with **dynamics** the Ising energy functions associated with problems amounts to run a **local search** algorithm

\[ P(\sigma) \propto e^{-\beta H(\sigma)} \]

Simulate the Ising models at decreasing temperatures (simulated annealing algorithms) to obtain approximate solutions for problems.
Computational problems

Endowing with **dynamics** the Ising energy functions associated with problems amounts to run a **local search** algorithm

$$P(\sigma) \propto e^{-\beta H(\sigma)}$$

Simulate the Ising models at decreasing temperatures (simulated annealing algorithms) to obtain approximate solutions for problems

Better understanding of algorithmic barriers of local algorithms
Computational problems

Random 3-SAT and other problems experience sharp transitions as some “density” is varied. Under-constrained to over-constrained transition

\[ \alpha = \frac{\# \text{ clauses}}{\# \text{ variables}} \]
Computational problems

Random 3-SAT and other problems experience sharp transitions as some “density” is varied. Under-constrained to over-constrained transition

\[ \alpha = \frac{\# \text{ clauses}}{\# \text{ variables}} \]

Easy-Hard-Easy?
Computational problems

Looking at scalability the picture seemed more like Easy-Hard-Less Hard
Computational problems

In 1990’s the relation between phase transitions and hardness was a big hype! Efforts to apply statistical physics to solve P/NP!
Computational problems

In 1990’s the relation between phase transitions and hardness was a big hype! Efforts to apply statistical physics to solve P/NP!

But:
Computational problems

In 1990’s the relation between phase transitions and hardness was a big hype! Efforts to apply statistical physics to solve P/NP!

But:

Counterexamples: random 2-SAT has a phase transition at density 1
Computational problems

In 1990's the relation between phase transitions and hardness was a big hype! Efforts to apply statistical physics to solve P/NP!

But:

Counterexamples: random 2-SAT has a phase transition at density 1

Average-case not worst-case complexity
Computational problems

In 1990’s the relation between phase transitions and hardness was a big hype! Efforts to apply statistical physics to solve P/NP!

But:

Counterexamples: random 2-SAT has a phase transition at density 1

Average-case not worst-case complexity

Phase transitions... will NOT solve the P/NP question
Computational problems
Computational problems

People started looking into the geometry of the space of solutions
Computational problems

People started looking into the geometry of the space of solutions

Random 4-SAT:
Computational problems

People started looking into the geometry of the space of solutions

Random 4-SAT:

**Adjacency:** Hamming distance 1

**Cluster:** connected component within the set of satisfying assignments
Computational problems

People started looking into the geometry of the space of solutions

Random 4-SAT:

**Adjacency:** Hamming distance 1

**Cluster:** connected component within the set of satisfying assignments
Computational problems

People started looking into the geometry of the space of solutions

Random 4-SAT:

Adjacency: Hamming distance 1

Cluster: connected component within the set of satisfying assignments

Clustering is conjectured to be a barrier for local algorithms!
Computational problems
Computational problems

Information Flow in a Kinetic Ising Model Peaks in the Disordered Phase
Phys. Rev. Lett. 111, 177203 – Published 24 October 2013
Lionel Barnett, Joseph T. Lizier, Michael Harré, Anil K. Seth, and Terry Bossomaier
Computational problems

**Research project:** apply information theoretic functionals (transfer entropy, information storage) to characterize the information exchange between nodes in the Ising graph associated to random 4-SAT
Outlook

Magnets & Ising model

Computational problems

Artificial neural networks

Retina
Machines (ANN)

Brain-inspired architecture for statistical learning that is used to approximate functions that depend on many inputs.

Units (neurons) are interconnected with adaptive weights, which sum up inputs and transform total input into output value.

+ Learn wild functions!
- Challenging to train
Deep learning in press...

Google Hires Brains that Helped Supercharge Machine Learning
Wired 3/2013

Facebook taps ‘Deep Learning’ Giant for new AI Lab
Wired 12/2013

Is “Deep Learning” A Revolution in Artificial Intelligence?
New Yorker 11/2012

Why Did Google Pay $400 Million for DeepMind
MIT Technology Reviews 1/2014

New Techniques from Google Are Taking Artificial Intelligence to Another Level
MIT Technology Reviews 5/2013
Google Scholar

“deep learning” + “neural network”
There is an algorithm to train the weights but...

...we need a good starting point!
Machines (ANN)

Boltzmann machines

**Goal:** generative model of data! Even in absence of input the dynamics will sample from the observed input distribution.
Machines (ANN)

Boltzmann machine

\[ E = - \left( \sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i \right) \]

\[ p_{i=\text{on}} = \frac{1}{1 + \exp(-\frac{\Delta E_i}{T})} \]

Ising (on a graph at T)

\[ H(\sigma) = - \sum_{\langle i, j \rangle} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i \]

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh\left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]
Machines (ANN)

Boltzmann machine

\[ E = - (\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i) \]

\[ p_{i=\text{on}} = \frac{1}{1 + \exp(-\Delta E_i / T)} \]

Ising (on a graph at \( T \))

\[ H(\sigma) = - \sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i \]

\[ f_i(\sigma) = \frac{1}{2} \left[ 1 - \tanh \left( \frac{\beta \Delta H_i(\sigma)}{2} \right) \right] \]

Once trained these machines are identical to Ising models!
Training a Boltzmann machine

\[ W' = W + \alpha (\langle vh \rangle_0 - \langle vh \rangle_k) \]

Contrastive divergence method changes the weights after \( k \) steps of the chain
Machines (ANN)

Stacked Boltzmann machines
Pre-training **ANN**: start with weights from a generative model such as a stacked RBM.
Machines (ANN)

That is how Ising-like models helped to classify cats from raw youtube videos
Machines (ANN)

Hopfield network

Fully-recurrent network with symmetric weights

Weights are determined before-hand by set of patterns to be memorized

Network starts with a corrupted or partially-complete pattern

Network dynamics causes pattern to be recalled

Each stored pattern acts as an attractor
Machines (ANN)

Hopfield network
Machines (ANN)

Hopfield network

\[
s_i \left\{ \begin{array}{ll}
'1' & \text{if } \sum_j w_{ij} s_j \geq \theta_i, \\
' - 1' & \text{otherwise.}
\end{array} \right.
\]

\[
E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i
\]

Ising (on a graph at T=0)

\[
H(\sigma) = -\sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i
\]
Outlook

Magnets and Ising model

Computational problems as Ising problems

Artificial neural networks as kinetic Ising

Retina and criticality
Brains (retina)
Vertebrate retina (salamander)
Vertebrate retina (salamander)

160 neurons simultaneously
Naturalistic stimulation
Vertebrate retina (salamander)

160 neurons simultaneously
Naturalistic stimulation
Brains (retina)

Aim: How well can an Ising model explain the firing data?

Brains (retina)

Strategy:

1. Define binary variables $\sigma_i$
2. Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3. Fit coefficients $J$ and $h$
Brains (retina)

**Strategy:**

1. Define binary variables $\sigma_i$
2. Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3. Fit coefficients $J$ and $h$

---

1. Discretize firing to define binary variables $\sigma_i$

   Bin size $\Delta t = 20 ms$

   Neuron $i$ fires 1 or more spikes in $\Delta t \rightarrow \sigma_i = 1$

   Neuron $i$ fires 0 spikes in $\Delta t \rightarrow \sigma_i = 0$
Brains (retina)

Strategy:

1 Define binary variables $\sigma_i$
2 Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3 Fit coefficients $J$ and $h$
Brains (retina)

Strategy:
1. Define binary variables $\sigma_i$
2. Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3. Fit coefficients J and h

2. Estimate measures from data

Mean probability of each neuron generating a spike $\langle \sigma_i \rangle$
Correlation between spiking in pairs of neurons $\langle \sigma_i \sigma_j \rangle$
Brains (retina)

Strategy:

1. Define binary variables $\sigma_i$
2. Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3. Fit coefficients $J$ and $h$
Strategy:

1. Define binary variables $\sigma_i$
2. Estimate $\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle$
3. Fit coefficients $J$ and $h$

3. Fit coefficients to match estimated measures

\[
P(\{\sigma_i\}) = \frac{1}{Z} \exp[-E(\{\sigma_i\})]
\]

\[
E(\{\sigma_i\}) = -\sum_{i=1}^{N} h_i \sigma_i - \frac{1}{2} \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j - V(K)
\]

Maximum entropy distribution up to second order moments
Brains (retina)

The resulting Ising model fits data statistics very well (even beyond the measures used for the fitting)!
Brains (retina)

The resulting Ising model fits data statistics very well (even beyond the measures used for the fitting)!

Changing the effective temperature of the Ising model... original fit (T=1) shows that retina is operating at a critical regime!
Brains (retina)

The resulting Ising model fits data statistics very well (even beyond the measures used for the fitting)!

Changing the effective temperature of the Ising model... original fit (T=1) shows that retina is operating at a critical regime!

Does retinal neurons adapts to operate at criticality?
Brains (retina)

The resulting Ising model fits data statistics very well (even beyond the measures used for the fitting)!

Changing the effective temperature of the Ising model... original fit ($T=1$) shows that retina is operating at a critical regime!

Does retinal neurons adapts to operate at criticality?

Simple reflection of power-law statistics of natural images?
Brains (retina)

Does the training of RBM under natural images result in criticality?

Research project:

1. Train RBM with a bunch of natural images
2. Obtain weights and biases (J’s and h’s)
3. Simulate resulting Ising model with different temperatures
4. Plot the heat capacity as a function of T
Conclusion and future work

Ising systems are canonical **models of collective phenomena**

Ising models **appear** in a wide **variety of contexts and problems**

**Exploit the understanding of Ising** models from decades of research in statistical physics **to improve understanding in other problems** in neuroscience, neural networks, and other fields

\[
H(\sigma) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i
\]

\[
P(\sigma) \propto e^{-\beta H(\sigma)}
\]

\[
\text{Ising models appear in a wide variety of contexts and problems}
\]

\[
\text{Exploit the understanding of Ising models from decades of research in statistical physics to improve understanding in other problems in neuroscience, neural networks, and other fields}
\]
Thanks!

Luiz Lana
Michael Wibral