Exponential Reduction in Sample Complexity with Learning of Ising Model Dynamics

Arkopal Dutt, Andrey Lokhov, **Marc Vuffray**, and Sidhant Misra

Motivation

Unsupervised learning task

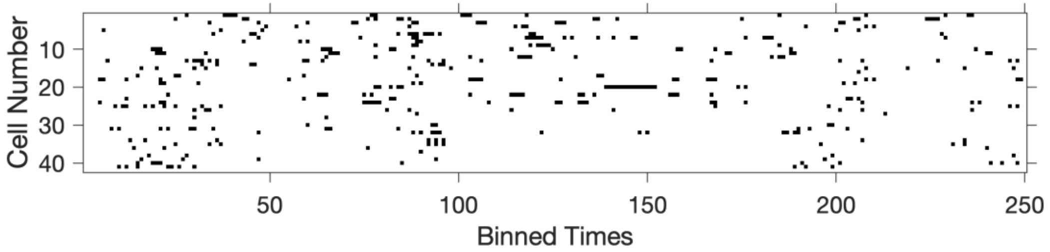
- Observe draws of random vectors $\underline{\sigma}$
- Learn structure and parameters of a positive distribution $\mu(\underline{\sigma}) > 0$

Limitation: assuming samples of $\underline{\sigma}$ are i.i.d.

Can we learn from non-i.i.d. time-correlated or dynamical samples and gain an advantage?

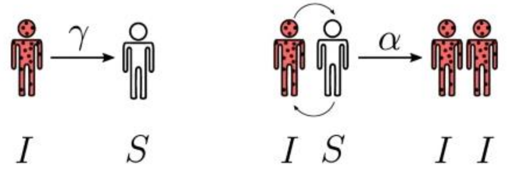
Applications/Scenarios

Neural Spike Trains[1]

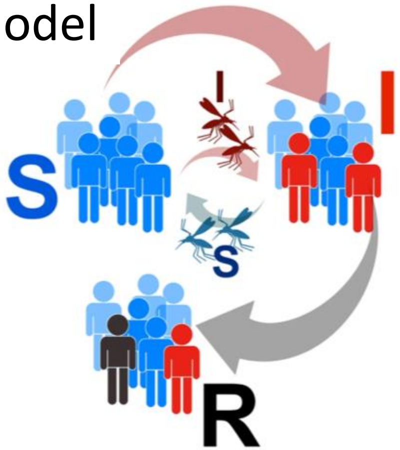


Epidemic Spreading[2,3]:

SIS Model



SIR Model

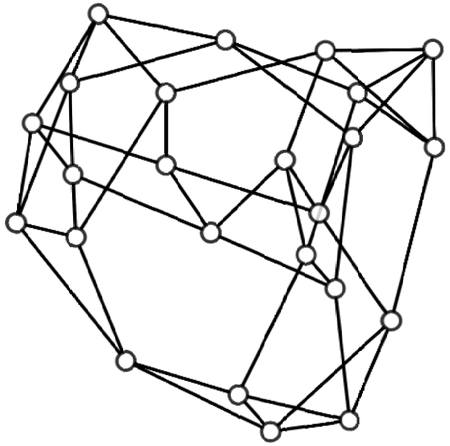


[1]Prentice et al., PLoS Computational Biology (2016); [2]Alfinito et al., Scientific Reports (2016); [3]Nakamura et al., Scientific Reports (2019)

Graphical Model Learning

Learn probability distribution $\mu(\underline{\sigma}) > 0$ which has conditional dependency structure according to a given graph *G=(V,E)*

Ising Models:



$$\mu(\underline{\sigma}) = \frac{1}{Z} \exp\left(\sum_{(i,j)\in E} J_{ij}\sigma_i\sigma_j + \sum_{i\in V} H_i\sigma_i \right)$$

- Binary RV at $i \in V$: $\sigma_i \in \{-1, 1\}$
- $\mu(\underline{\sigma})$ parameterized by
  - Coupling intensity $\underline{J} = \{J_{ij} | (i,j) \in E\}$
  - Magnetic field $\underline{H} = \{H_i | i \in V\}$
- Naturally defined dynamics
- Properties:
  - node degree d
  - maximum coupling intensity β

Source: Ghahramani, Z., Journal of Pattern Recognition and AI (2001)

Ising Model Learning

Structure learning: learn $\{(i,j) | J_{ij} \neq 0\}$

Parameter learning: learn $\underline{J}$ and $\underline{H}$ up to some accuracy

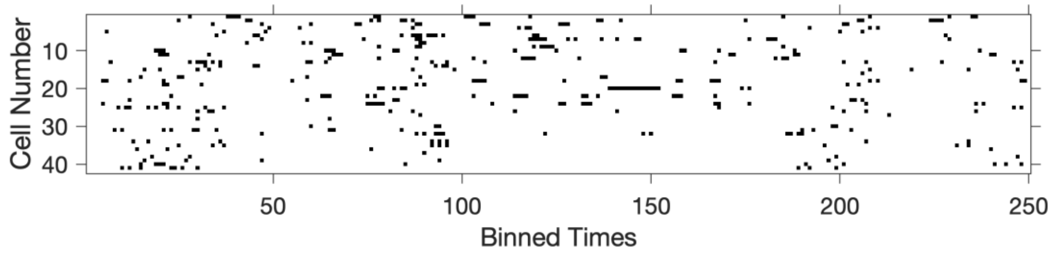
Sample complexity[1] (i.i.d.): $m = \mathcal{O}\left(\exp(6\beta d)\right)$

Can we learn Ising models efficiently from time-correlated samples?

[1]Vuffray, Misra, Lokhov (NeurIPS 2016, Science Advances 2018)

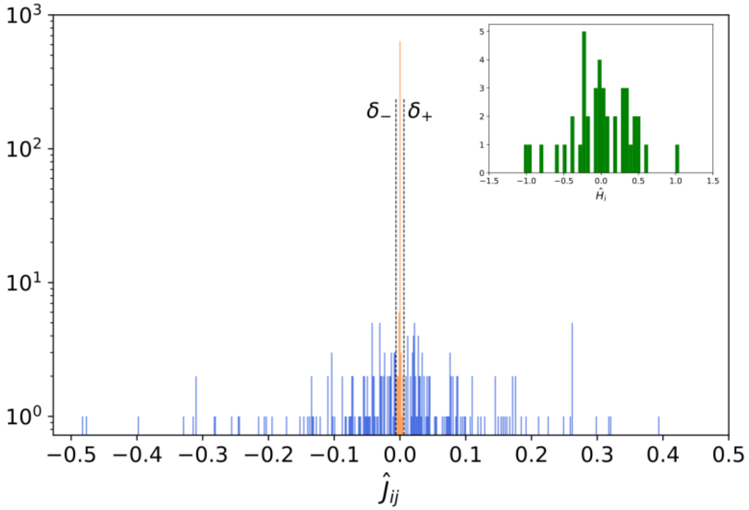
Dynamical Learning for Biological Neuronal Networks

Neural Spike Trains



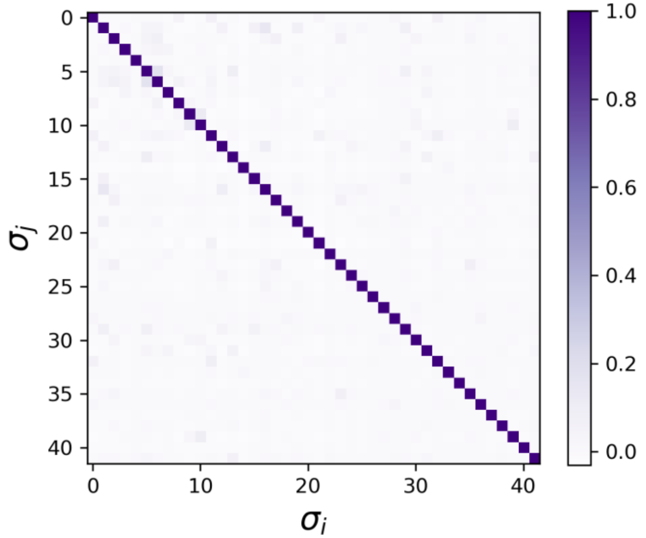
Dynamical Learning

Graphical Model Parameters

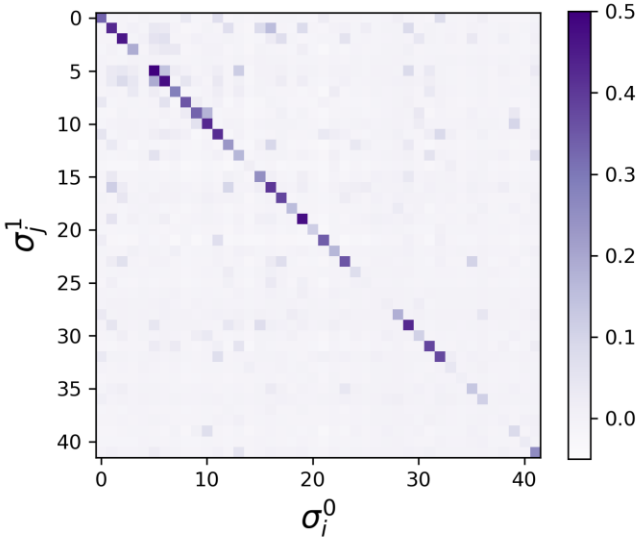


Correlations:

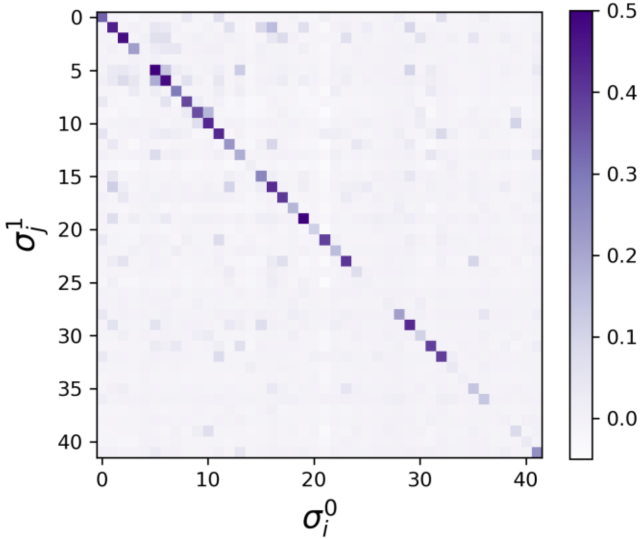
Data assumed to be i.i.d.



Data respecting dynamics



Model predictions



Generating samples through dynamics

Natural dynamics on Ising models: Glauber dynamics (Gibbs sampling)

At time step *t*:

$$\underline{\sigma}^t \longrightarrow \underline{\sigma}^{t+1}$$

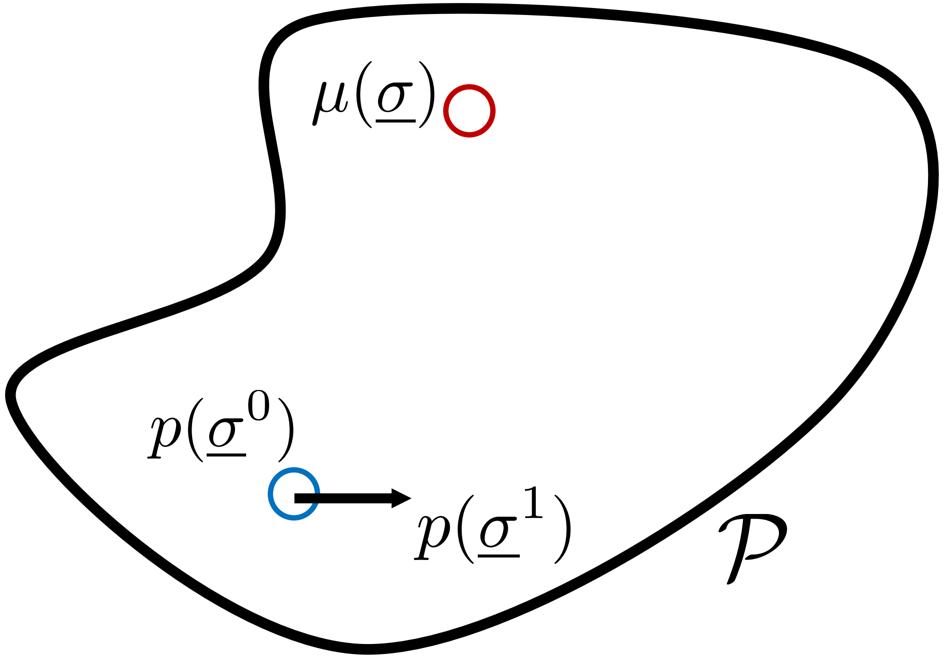
Choose node i at random

Update $\sigma_i^{t+1} \sim p(\sigma_i | \underline{\sigma}^t)$

Sample: $(\underline{\sigma}^t, \underline{\sigma}^{t+1}, \underbrace{I^{t+1}}_{\text{Updated node identity}})$

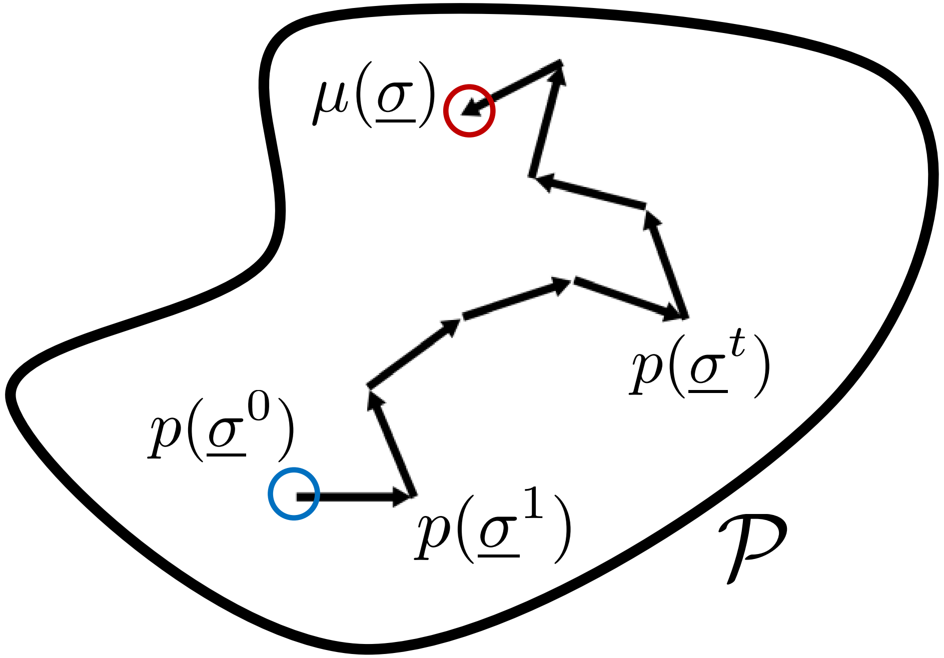
Setting of Learning Ising Models from Dynamics

T-Regime (Trajectory)



Setting of Learning Ising Models from Dynamics

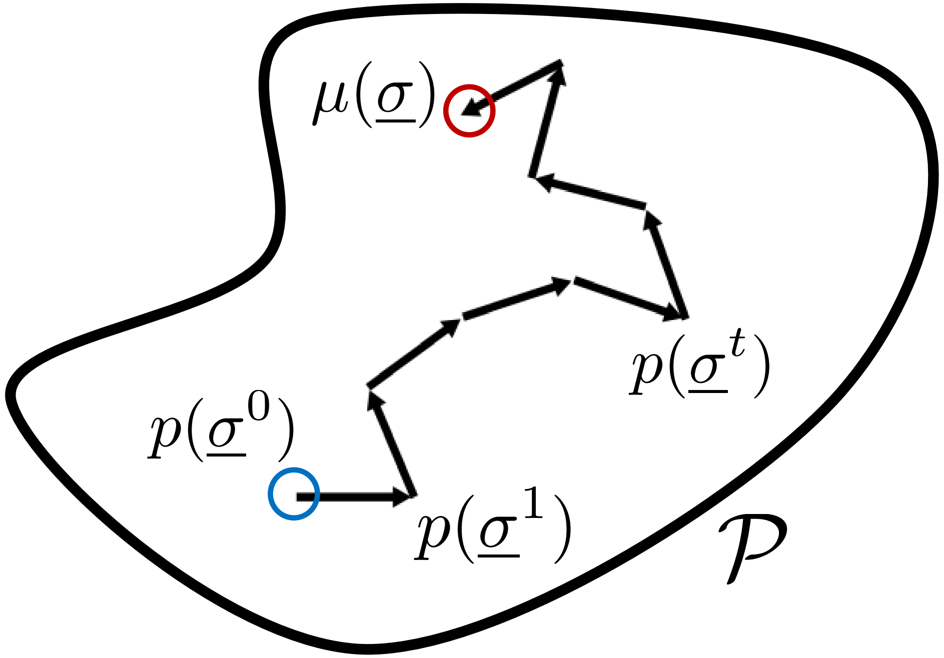
T-Regime (Trajectory)



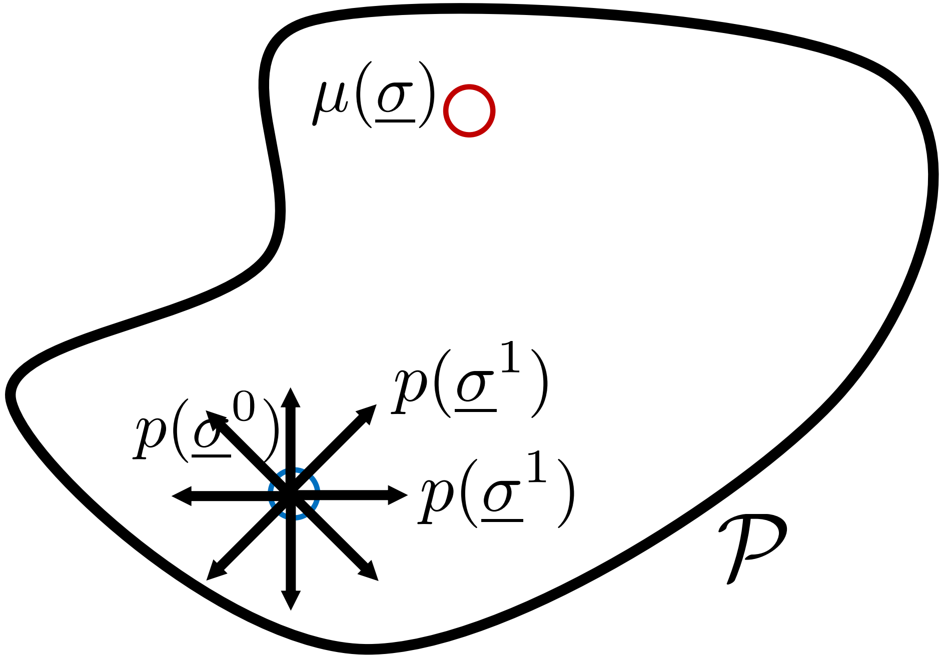
- Initial distribution $p(\underline{\sigma}^0)$ is uniform distribution
- Mixing time is exponential in β

Setting of Learning Ising Models from Dynamics

T-Regime (Trajectory)



M-Regime (Multiple Restarts)



- Initial distribution $p(\underline{\sigma}^0)$ is uniform distribution
- Mixing time is exponential in β

Sampling far from equilibrium

Efficient Algorithms for Learning Ising Model Dynamics

Adapted learning algorithms[1,2] from i.i.d. samples to Glauber Dynamics

Input: *m* samples $\{(\underline{\sigma}^t, \underline{\sigma}^{t+1}, I^{t+1})\}_{t \in \{0,1,\ldots,m-1\}}$

Idea: For each node $u \in V$, maximize conditional likelihood (Glauber dynamics)

$$p(\sigma_u^{t+1} | \underline{\sigma}^t)$$

[1]Ravikumar et al. (2010); [2]Vuffray et al. (2016), Lokhov et al. (2018)

Efficient Algorithms for Learning Ising Model Dynamics

Form: $$(\underline{\hat{J}}_u, \hat{H}_u) = \underset{(\underline{J}_u, H_u)}{\operatorname{argmin}} \mathcal{L}_m(\underline{J}_u, H_u) + \lambda ||\underline{J}_u||_1$$

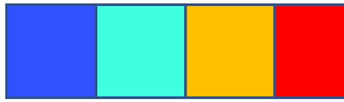
Estimators:

- Dynamics Regularized Pseudolikelihood Estimation (D-RPLE)
- Dynamics Regularized Interaction Screening Estimation (D-RISE)

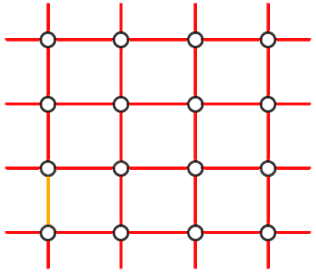
Local Reconstruction (one neighborhood at a time)

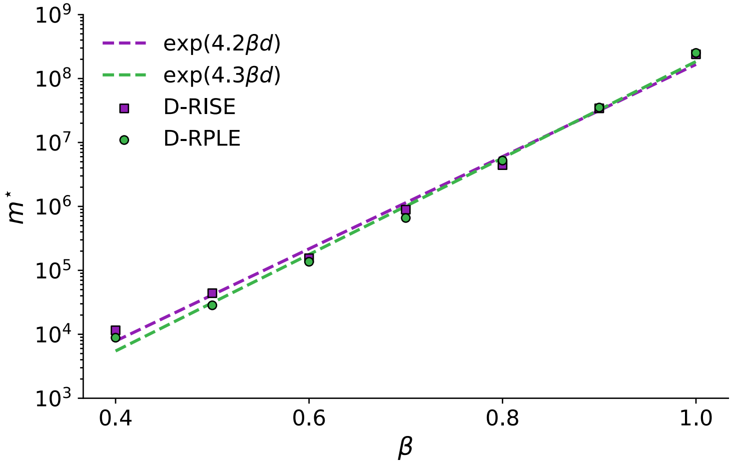
Convex Function (with low computational complexity e.g., using entropic descent)

Empirical Study of Sample Complexity in T-regime

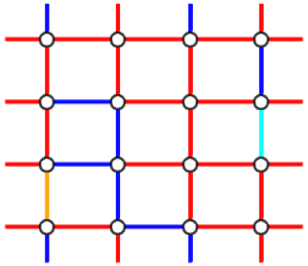


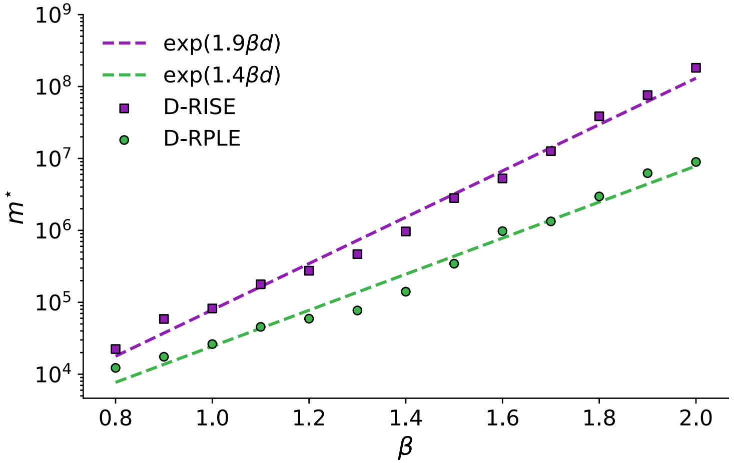
Ferromagnetic model



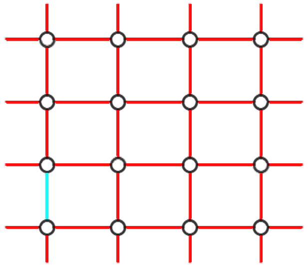


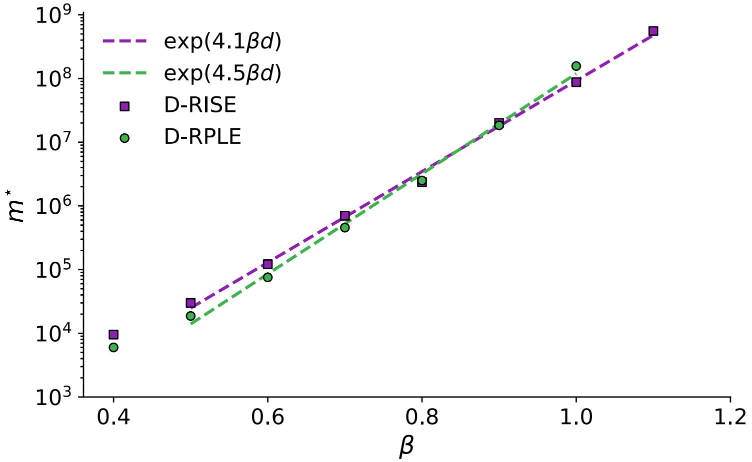
Spin glass





Ferromagnetic model with weak impurity

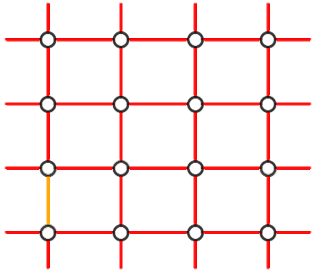


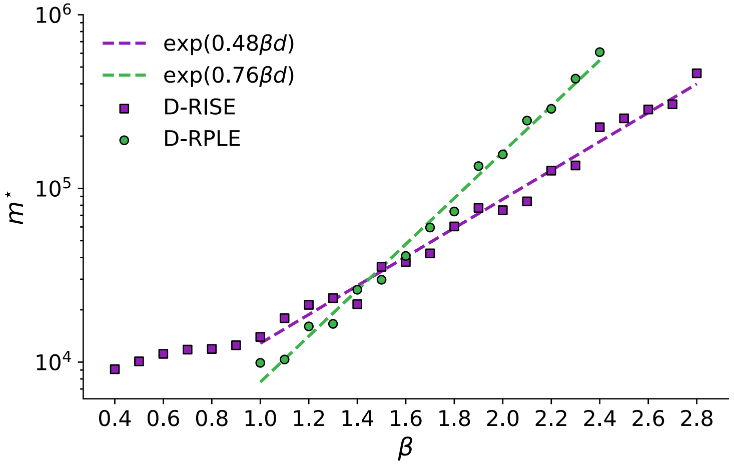


Sample complexity similar to i.i.d. case (well-mixed)

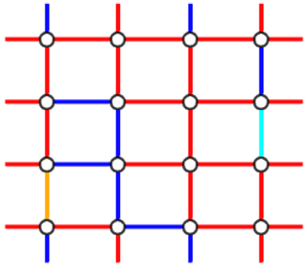
Empirical Study of Sample Complexity in M-regime

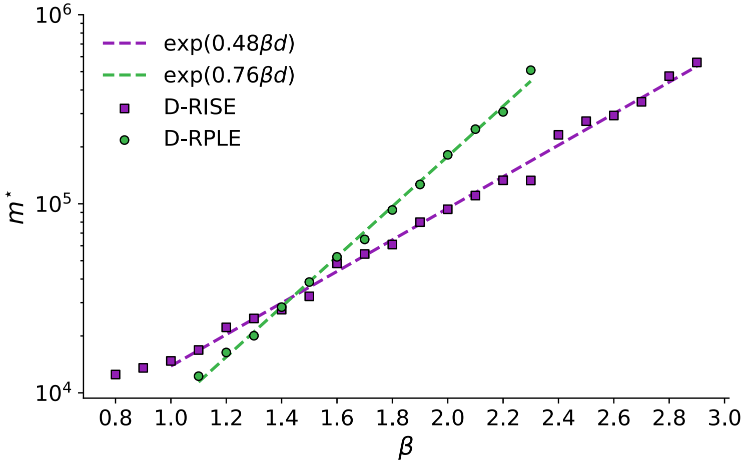
Ferromagnetic model



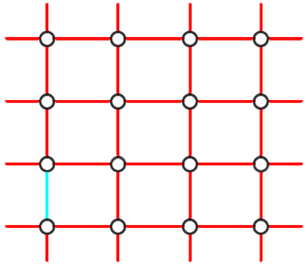


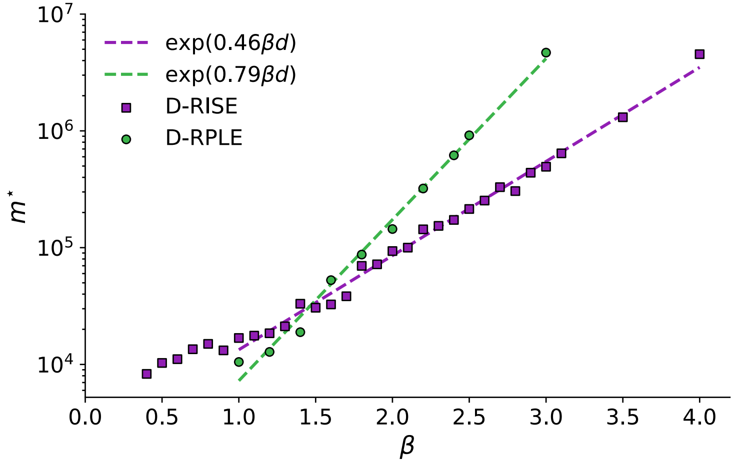
Spin glass





Ferromagnetic model with weak impurity





Exponential reduction in sample complexity compared to i.i.d. or T-regime

Theorem for Learning of Ising Models in M-regime

Informal: With high probability, learning algorithms learn the parameters accurately for all nodes $u \in V$, if the number of samples satisfy

D-RPLE: $m = \mathcal{O}\left(\exp(4\beta d)\right)$

D-RISE: $m = \mathcal{O}\left(\exp(2\beta d)\right)$

Ising model specific properties

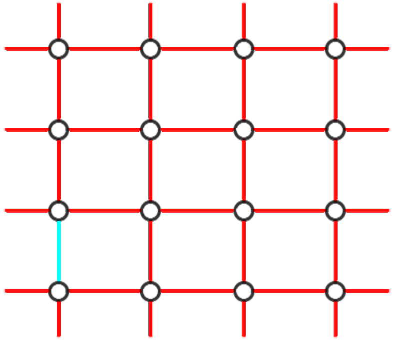
- Node degree: d
- Maximum coupling intensity: β

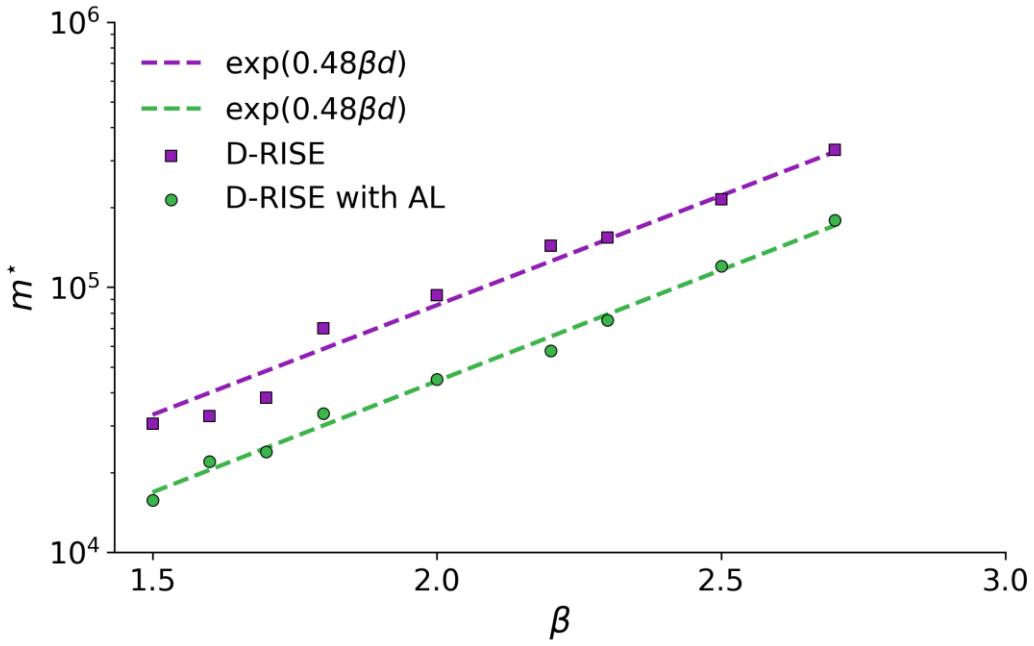
Active Learning of Ising Model Dynamics

Can we improve the sample complexity of learning Ising models in M-regime through a wise choice of initial query distribution $p(\underline{\sigma}^0)$?

- M-regime is amenable to both online and active learning
- *Max-entropy distribution* yields up to 47% constant savings

Ferromagnetic model with weak impurity





Summary

Results and Implications

- Time correlated samples can be useful for unsupervised learning
- Ising models can be efficiently learned from Glauber dynamics
- Highlighted real-world applications

Future Work

- Extension to multi-site dynamics, MRFs, partial observations, etc.