Deep Boltzmann Machine: on the annealed and the replica symmetric region

Emanuele Mingione
joint work with D.Alberici, A. Barra, P. Contucci

emanuele.mingione2@unibo.it

Università di Bologna

arXiv: 2004.04495
arXiv: 2001.07714

Outline

- Definition of DBM and basic results
- Annealed region
- Replica symmetric bound

Related works on bipartite models:

A. Barra, G. Genovese, F. Guerra,(2011)
Auffinger, W.-K. Chen, (2014)
J. Baik, J.O. Lee, (2017)
J. Barbier, N. Macris, L. Miolane, (2017)

Deep Boltzmann Machine

Deep (restricted) Boltzmann Machines (DBM) are generative models, first introduced in by Salakhutdinov-Hinton (2009)

Main feature: deep architecture (linear chain of Boltzmann Machines)

We look at this model from point of view of spin glass theory (quenched setting)

Main results: complete characterization of the annelead region, replica symmetric bound

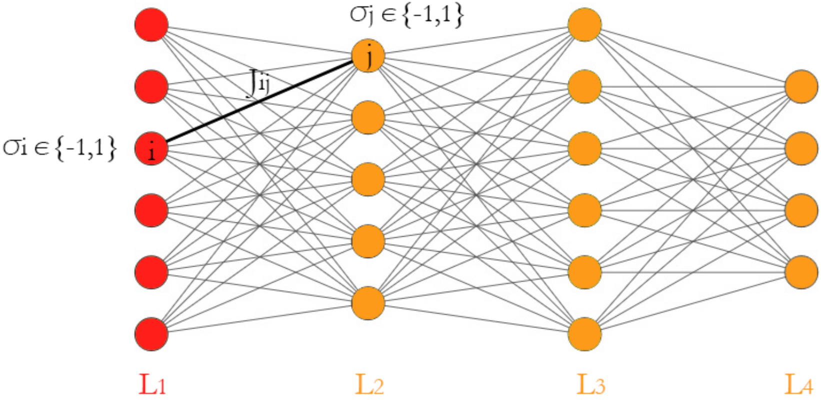
Deep Boltzmann Machine

N binary variables (spins, neurons...) are distributed along K layers $L_1, \ldots, L_K$ with $|L_p| = N_p$ for $p = 1, \ldots, K$ and $\sum_{p=1}^{K} N_p = N$:

$$\sigma_i \in \{-1, 1\} \quad \text{for } i \in L_1 \cup \cdots \cup L_K \equiv \Lambda_N \, .$$

- we assume $\frac{N_p}{N} \xrightarrow[N\to\infty]{} \lambda_p \in [0,1]$ for every $p = 1, \ldots, K$. We denote $\lambda = (\lambda_p)_{p=1,\ldots,K}$. Clearly $\sum_{p=1}^{K} \lambda_p = 1$.
- J_{ij} for $(i,j) \in L_p \times L_{p+1}$ and $p = 1, \ldots, K-1$ is a family of i.i.d. standard Gaussian random variables coupling spins in two consecutive layers
- h_i for $i \in L_p$ and $p = 1, \ldots, K$ be i.i.d. copies of a random variable $h^{(p)}$ such that $\mathbb{E}|h^{(p)}| < \infty$. We denote $h = (h^{(p)})_{p=1,\ldots,K}$.
- $\beta = (\beta_p)_{p=1,\ldots,K-1} \in \mathbb{R}_+^{K-1}$ is a vector of positive "inverse temperatures" tuning the interactions among consecutive layers

Deep Boltzmann Machine



Hamiltonian

Definition

The Hamiltonian of the random Deep Boltzmann Machine [DBM] is

$$H_{\Lambda_N}(\sigma) \equiv -\frac{\sqrt{2}}{\sqrt{N}} \sum_{p=1}^{K-1} \beta_p \sum_{(i,j)\in L_p \times L_{p+1}} J_{ij}\, \sigma_i \sigma_j \tag{1}$$

for every spin configuration $\sigma \in \{-1,1\}^N$.

Definition

Given two spin configurations $\sigma, \tau \in \{-1,1\}^N$, for every $p = 1, \dots, K$ we define their overlap over the layer L_p as

$$q_{L_p}(\sigma,\tau) \equiv \frac{1}{N_p} \sum_{i \in L_p} \sigma_i\, \tau_i \ \in [-1,1]\,. \tag{2}$$

Covariance matrix

The covariance matrix of the centred Gaussian process H_{Λ_N} is

$$\mathbb{E}\, H_{\Lambda_N}(\sigma)\, H_{\Lambda_N}(\tau) \;=\; N\, q_{\Lambda_N}(\sigma,\tau)^T\, M_1\, q_{\Lambda_N}(\sigma,\tau) \tag{3}$$

for every $\sigma, \tau \in \{-1, 1\}^N$. Here we set $q_{\Lambda_N}(\sigma,\tau) \equiv \big(q_{L_p}(\sigma,\tau)\big)_{p=1,\dots,K}$,

$$M_1(\beta,\lambda) \;\equiv\; \mathrm{diag}(\lambda)\, M_0(\beta)\; \mathrm{diag}(\lambda)\ , \tag{4}$$

$$M_0(\beta) \;\equiv\; \begin{pmatrix} 0 & \beta_1^2 & & & & \\ \beta_1^2 & 0 & \beta_2^2 & & & \\ & \beta_2^2 & 0 & & & \\ & & & \ddots & & \\ & & & & & \beta_{K-1}^2 \\ & & & & \beta_{K-1}^2 & 0 \end{pmatrix} \tag{5}$$

Notice that $M_0(\beta)$ can be interpreted as a weighted adjacency matrix for the layers structure of the DBM.

Quenched Pressure

Definition

The random partition function of the model introduced by Hamiltonian (1) is

$$Z_{\Lambda_N} \equiv \sum_{\sigma\in\{-1,1\}^N} \exp\left(- H_{\Lambda_N}(\sigma) + \sum_{p=1}^{K} \sum_{i\in L_p} h_i\,\sigma_i \right) \tag{6}$$

and its quenched pressure density is

$$\mathrm{p}^{\mathrm{DBM}}_{\Lambda_\mathrm{N}} \equiv \frac{1}{N}\,\mathbb{E}\,\log Z_{\Lambda_N} \tag{7}$$

where $\mathbb{E}$ denotes the expectation over all the couplings J_{ij} 's and the external fields h_i's.

A lower bound for p^{DBM}

For every $a = (a_p)_{p=1,\dots,K-1} \in \mathbb{R}_+^{K-1}$ we define

$$\mathcal{P}^{DBM}(a) \equiv \sum_{p=1}^{K} \lambda_p \, p_N^{SK}\left(\theta_p(a), h^{(p)}\right) - \frac{1}{2} \sum_{p=1}^{K} \lambda_p \, \theta_p(a)^2 + \sum_{p=1}^{K-1} \lambda_p \, \beta_p^2 \, \lambda_{p+1} \tag{8}$$

where $p_N^{SK}(\beta, h)$ is a quenched pressure of an Sherringhton-Kirkpatrick model of N particle at inverse temperature β and external field h and $\theta_p(a) = \theta_p(a; \beta, \lambda) \geq 0$ is defined by:

$$\theta_p(a)^2 \equiv \begin{cases} \lambda_1 \, a_1 \, \beta_1^2 & \text{for } p = 1 \\ \lambda_p \left(\dfrac{1}{a_{p-1}} \beta_{p-1}^2 + a_p \, \beta_p^2 \right) & \text{for } p = 2, \dots, K-1 \\ \lambda_K \, \dfrac{1}{a_{K-1}} \beta_{K-1}^2 & \text{for } p = K \end{cases} . \tag{9}$$

A lower bound for $\mathrm{p}^{\mathrm{DBM}}$

Theorem

The quenched pressure of the DBM satisfies the following lower bound:

$$\mathrm{p}^{\mathrm{DBM}}_{\Lambda_{\mathrm{N}}} \geq \mathcal{P}^{\mathrm{DBM}}_{\mathrm{N}}(a) \;, \tag{10}$$

for every $a = (a_p)_{p=1,\ldots,K-1} \in \mathbb{R}_+^{K-1}$.

Proof idea: interpolate between the DBM with K layers and K Sherringhton-Kirkpatric models: the bound folllows from the inequality

$$\left(a\, q_{L_p} - \frac{1}{a}\, q_{L_{p+1}} \right)^2 \geq 0, \quad \forall\, a > 0$$

The annealed region

For zero external field ($h = 0$) using the above theorem one can identify a region where the quenched and the annealed pressure of the DBM coincide.

Definition

The annealed pressure of the DBM is

$$\mathrm{p}^{\text{DBM-A}} \equiv \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\, Z_{\Lambda_N} \,. \tag{11}$$

It can be easily computed due to the Gaussian nature of the model:

$$\mathrm{p}^{\text{DBM-A}}(\beta, \lambda) = \log 2 + \sum_{p=1}^{K-1} \lambda_p\, \beta_p^2\, \lambda_{p+1} \,. \tag{12}$$

By concavity of the log, the annealed pressure is an upper bound for the quenched one:

$$\limsup_{N\to\infty} \mathrm{p}^{\text{DBM}}_{\Lambda_\text{N}} \leq \mathrm{p}^{\text{DBM-A}} \,. \tag{13}$$

The system is said to be in the *annealed regime* when the parameters (β, λ) are such that $\lim_{N\to\infty} \mathrm{p}^{\text{DBM}}_{\Lambda_\text{N}} = \mathrm{p}^{\text{DBM-A}}$.

The annealed region

The annealed regime of the DBM can be idetinfied exploiting the knowledge of the annealed regime of the SK model.

Let p^{SK} be the limiting quenched pressure of an SK model and let $\mathrm{p}^{\mathrm{SK\text{-}A}} \equiv \lim_{N\to\infty} N^{-1} \log \mathbb{E} Z_N^{\mathrm{SK}}$ be its annealed version. By Jensen inequaslity

$$\mathrm{p}^{\mathrm{SK}} \le \mathrm{p}^{\mathrm{SK\text{-}A}} = \log 2 + \frac{\beta^2}{2} \,. \tag{14}$$

Equality is achieved in the annealed region of the SK model

$$\mathrm{p}^{\mathrm{SK}}(\beta) = \mathrm{p}^{\mathrm{SK\text{-}A}}(\beta) \quad \text{if } \beta^2 \le \frac{1}{2} \,. \tag{15}$$

The annealed region

Consider the following system of inequalities:

$$\begin{cases} \lambda_1 \, a_1 \, \beta_1^2 < \dfrac{1}{2} \\ \lambda_p \left(\dfrac{1}{a_{p-1}} \, \beta_{p-1}^2 + a_p \, \beta_p^2 \right) < \dfrac{1}{2} & \text{for } p = 2, \ldots, K-1 \\ \lambda_K \, \dfrac{1}{a_{K-1}} \, \beta_{K-1}^2 < \dfrac{1}{2} \end{cases} \tag{16}$$

and define

$$A_K \equiv \left\{ (\beta, \lambda) \in \mathbb{R}_+^{K-1} \times T_K \;\middle|\; \exists \, a \in \mathbb{R}_+^{K-1} : (16) \text{ is verified} \right\} , \tag{17}$$

where $T_K \equiv \{(\lambda_1, \ldots, \lambda_K) \in [0,1]^K \mid \sum_{p=1}^K \lambda_p = 1\}$ denotes the $K-$dimensional simplex. We denote by $\overline{A_K}$ the topological closure of A_K .

The annealed region

Theorem

If $(\beta, \lambda) \in \overline{A_K}$ *there exists*

$$\lim_{N\to\infty} \mathrm{p}^{\mathrm{DBM}}_{\Lambda_{\mathrm{N}}} = \mathrm{p}^{\mathrm{DBM\text{-}A}} \,. \tag{18}$$

The region A_K is given in terms of implicit conditions on β, λ however there exists a mapping between A_K and matching polynomials that is useful to investigste in more detail the annealed region

The annealed region and matching polynomials

Definition

Let $x \in \mathbb{R}$ and $t = (t_p)_{p=1,\dots,K-1} \in [0,\infty)^{K-1}$. We define recursively

$$\begin{cases} \Delta_{p+1}(x,t) \equiv x\ \Delta_p(x,t) - t_p\ \Delta_{p-1}(x,t) \quad \text{for } p = 1,\dots,K-1 \\ \Delta_1(x,t) \equiv x\,,\ \Delta_0(x,t) \equiv 1 \end{cases} . \tag{19}$$

These polynomials have several characterizations and were studied by Heilmann and Lieb in the context of monomer dimer models.

Let $(\beta,\lambda)\in\mathbb{R}_+^{K-1}\times T_K$, consider the vector $t=(t_p)_{p=1,\dots,K-1}$ with

$$t_p(\beta,\lambda)\;\equiv\;4\,\lambda_p\,\beta_p^4\,\lambda_{p+1} \tag{20}$$

for every $p=1,\dots,K-1$. Define

$$\rho(\beta,\lambda)\;\equiv\;\max\left\{x>0:\;\Delta_K\big(x,\;t(\beta,\lambda)\big)=0\right\}\,. \tag{21}$$

The followings are equivalent:

i) $(\beta,\lambda)\in A_K$;

ii) $\Delta_p\big(1,\;t(\beta,\lambda)\big)>0$ for every $p=2,\dots,K$;

iii) $\rho(\beta,\lambda)<1$.

A replica symmetric bound

The main theorem can be used to obtain a lower bound for the quenched pressure of the DBM in terms of the replica symmetric functional in a **suitable region** of the parameters β, λ, h.

For centred Gaussian external fields this region is defined though a system of K inequalities which mimic the Almeida-Thouless condition for the SK model.

Replica symmetric solution for the SK model

We denote by $\mathcal{P}^{\text{RS-SK}}$ the replica symmetric functional of the SK model, namely for every $q \in [0,1]$, $\beta > 0$, h real random variable with $\mathbb{E}\,|h| < \infty$,

$$\mathcal{P}^{\text{RS-SK}}(q;\, \beta, h) \,\equiv\, \mathbb{E} \log\cosh\left(z\,\sqrt{2\,q\,\beta^2} + h\right) + \frac{\beta^2}{2}\,(1-q)^2 + \log 2 \tag{22}$$

where z is a standard Gaussian random variable independent of h. Stationary points of $\mathcal{P}^{\text{RS-SK}}$ are identified by the consistency equation

$$q \,=\, \mathbb{E}\tanh^2\left(z\,\sqrt{2\,q\,\beta^2} + h\right) \tag{23}$$

where z is a standard Gaussian r.v. independent of h. The celebrated Guerra's bound states in particular that

$$\mathrm{p}^{\text{SK}}(\beta, h) \,\leq\, \inf_q \mathcal{P}^{\text{RS-SK}}(q;\, \beta, h)\,. \tag{24}$$

for every β, h. Identifying the exact replica symmetric region of the SK model, where equality in (24) is achieved, is an open problem.

The replica symmetric solution for the DBM

Definition

For $q = (q_p)_{p=1,\dots,K} \in [0,1]^K$ the *replica symmetric functional* of the DBM is

$$\mathcal{P}^{\text{RS-DBM}}(q;\,\beta,\lambda,h) \equiv \sum_{p=1}^{K} \lambda_p \, \mathbb{E} \log\cosh \left(z \sqrt{(Mq)_p} + h^{(p)} \right) + \\ + \frac{1}{2} (1-q)^T M_1 (1-q) + \log 2 \tag{25}$$

where $M = 2M_0 \,\mathrm{diag}(\lambda)$ and M_1 are tridiagonal matrices.

The stationary condition is

$$q_p = \mathbb{E} \tanh^2 \left(z \sqrt{(Mq)_p} + h^{(p)} \right) \quad \forall\, p = 1, \dots, K \,. \tag{26}$$

A replica symmetric bound

A first result about the replica symmetric region of the DBM under general (but implicit) conditions is provided by the following

Theorem

For β, λ, h such that there exist $q \in [0,1]^K$ and $a \in \mathbb{R}_+^{K-1}$ with

$$\lambda_p \, q_p \, a_p \; = \; \lambda_{p+1} \, q_{p+1} \quad \forall \, p = 1, \ldots, K-1 \tag{27}$$

and verifying

$$\mathrm{p}^{\mathrm{SK}} \left(\theta_p(a), h^{(p)} \right) \; = \; \mathcal{P}^{\text{RS-SK}} \left(q_p \, ; \theta_p(a), h^{(p)} \right) \quad \forall \, p = 1, \ldots, K \; , \tag{28}$$

then

$$\liminf_{N \to \infty} \mathrm{p}_{\Lambda_N}^{\mathrm{DBM}} \; \geq \; \mathcal{P}^{\text{RS-DBM}}(q; \, \beta, \lambda, h) \; . \tag{29}$$

A replica symmetric bound

More explicit conditions on β, λ, h for the replica symmetric bound (29) can be obtained trough the control of the replica symmetric region in the SK model: For β small enough Talagrand proved that for every h

$$\mathrm{p}^{\mathrm{SK}}(\beta, h) = \mathcal{P}^{\text{RS-SK}}(q;\, \beta, h) \quad \text{if } \beta^2 < \frac{1}{8} \tag{30}$$

where q is the unique solution of (23)

Corollary

Let β, λ, h such that a solution q of the replica symmetric consistency equation (26) *satisfies the inequalities*

$$(Mq)_p < \frac{1}{4}\, q_p \quad \forall\, p = 1, \ldots, K \tag{31}$$

Then the replica symmetric bound (29) *holds true.*

A replica symmetric bound

A necessary condition for replica symmetry on the SK model is the Almeida-Thouless condition:

$$\beta^2 \, \mathbb{E} \cosh^{-4}\left(z\sqrt{2\,q\,\beta^2} + h\right) \le \frac{1}{2} \tag{32}$$

where q is a solution of the consistency equation (23). If we take h Gaussian centered r.v. with variance $v > 0$, it was recently proved by W.K. Chen that the AT condition is also sufficient

Corollary

Assume $h^{(p)}$, $p = 1, \ldots, K$ centered Gaussian variables of variance $v_p > 0$ respectively. Let β, λ, v such that the (unique) solution q of the replica symmetric consistency equation (26) satisfies the inequalities

$$(Mq)_p \; \mathbb{E} \cosh^{-4}\left(z\sqrt{(Mq)_p + v_p}\right) \le q_p \qquad \forall\, p = 1, \ldots, K\,. \tag{33}$$

Then the replica symmetric bound (29) holds true.

THANK YOU!