IV Generative models

let's recall : I GMM unsupervised clustering
II, III: Perceptron . supervised classif
AE

A interesting task: to learn the data distribution
& to be able to generate $\boxed{\text{new data}}$
according to this dist.

concrete eff: in health with the privacy problem.
biology → genetics

A) Ising inverse problem

Imagine that you observe a set of N binary variables : $\{S_i^{(a)}\}_{i=1,\dots,N}^{a=1,\dots,M}$

± 1
0,1   2 ex: . neuron spikes
. b & w images

N spins
M samples

What would be the "simplest" model capable of reproducing (largest possible entropy)

- pairwise correlations : $\langle s_i s_j \rangle_D = \frac{1}{M} \sum_a s_i^a s_j^a$
- single site frequency : $\langle s_i \rangle_D = \frac{1}{M} \sum_a s_i^a$

Let's look at maximum entropy formulation

find $p(\{\vec{s}\})$ s.t.

$$S = -\sum_{\{s\}} p(\{\vec{s}\}) \log p(\{\vec{s}\}) + \sum_{i<j} \lambda_{ij} \Big(\overbrace{\sum_{\{s\}} s_i s_j \, p(\{\vec{s}\})}^{\langle s_i s_j \rangle_p} - \langle s_i s_j \rangle_D \Big)$$

$$+ \sum_i \lambda_i \left(\langle s_i \rangle_p - \langle s_i \rangle_D \right) + \lambda \Big(\sum_{\{s\}} p(\{\vec{s}\}) - 1 \Big)$$

$\left\{ \{\lambda_{ij}\}, \{\lambda_i\}, \lambda \right\}$: lagrange multipliers

$$\frac{\partial S}{\partial \lambda} = 0 \Rightarrow \sum p(\{\vec{s}\}) = 1 \;; \quad \frac{\partial S}{\partial \lambda_{ij}} = 0 \Leftrightarrow \langle s_i s_j \rangle_D = \langle s_i s_j \rangle_p \,, \quad \frac{\partial S}{\partial \lambda_i} = 0 \Leftrightarrow \langle s_i \rangle_p = \langle s_i \rangle_D$$

max over p : $$\frac{\delta S}{\delta p(\{\vec{s}\})} = 0 \Leftrightarrow -\log p(\{\vec{s}\}) - 1 + \sum_{i<j} s_i s_j \lambda_{ij} + \sum_i s_i \lambda_i + \lambda = 0$$

$$\Rightarrow p(\{\vec{s}\}) = \frac{e^{\sum_{i<j} \lambda_{ij} s_i s_j + \sum_i \lambda_i s_i}}{e^{1-\lambda}}$$

$\oplus$ norm. of p $\Rightarrow$ $$e^{1-\lambda} = Z = \sum_{\{\vec{s}\}} e^{\sum_{i<j} \lambda_{ij} s_i s_j + \sum_i \lambda_i s_i}$$

$$p(\{\vec{s}\}) = \frac{1}{Z} \exp\Big[\sum_{i<j} \lambda_{ij} s_i s_j + \sum_i \lambda_i s_i \Big] \quad : \text{Ising model.}$$

Hence: we need to find $\{\lambda_{ij}, \lambda_i\}$ such that the correlation and single site average matches the one of your dataset!

average matches the one of your dataset.

How? maximum-likelihood.

$$\{\lambda_{ij}, \lambda_i\} = \text{argmax}\left[\sum_a \log p(\{\vec{s}^{(a)}\})\right]$$

impossible to compute $O(2^N)$

$$= \text{argmax}\left\{ M\left[\underbrace{\left(\sum_{i<j} \lambda_{ij}\langle s_i s_j\rangle_D - \sum_i \lambda_i \langle s_i\rangle_D\right) - \log Z}_{\mathcal{L}}\right]\right\}$$

We will make a gradient ascent.

$$\frac{\partial \mathcal{L}}{\partial \lambda_{ij}} = \langle s_i s_j\rangle_D - \frac{\partial \log Z}{\partial \lambda_{ij}} = \langle s_i s_j\rangle_D - \frac{\sum_{\{s\}} s_i s_j e^{(\dots)}}{Z} = \langle s_i s_j\rangle_D - \langle s_i s_j\rangle_p$$

can be done using MC simu

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \langle s_i\rangle_D - \langle s_i\rangle_p$$

i) Good news: $\mathcal{L}$ is convex

ii) Bad news: $\langle\ \rangle_p$: hard to compute

more pbs:
- you only reproduce PW & fq., what higher order stat?
- lack of patterns

to address this pb

B] The Hopfield model

We define a set of P patterns: $\left\{\xi_i^{(\mu)} = \pm 1\right\}$ $\mu = 1, \dots, P$; $i = 1, \dots, N$

$\{\xi_i = \pm 1\}$ $i = 1, \dots, N$

with equal probability $\left[\begin{array}{l} p(+1) = 1/2 \\ p(-1) = 1/2 \end{array}\right.$ $\longrightarrow \sum_i \xi_i^\mu \xi_i^\nu \underset{N\to\infty}{=} \delta_{\mu\nu}$

Hopfield's Hamiltonian

$$\mathcal{H} = -\frac{1}{N}\sum_\mu \sum_{i<j} \xi_i^\mu \xi_j^\mu \underbrace{S_i S_j}_{\text{Ising spin } \pm 1} = -\frac{1}{N}\sum_{i<j} \underbrace{\left(\sum_\mu \xi_i^\mu \xi_j^\mu\right)}_{J_{ij}} S_i S_j$$

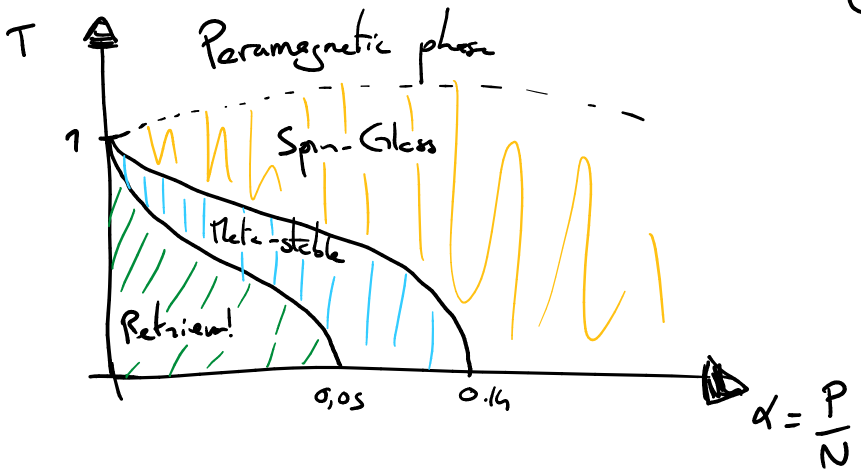
$$= -\frac{1}{2N}\sum_\mu \left(\sum_i \xi_i^\mu S_i\right)^2$$

remark: if $P = 1$, I can make the transformation $\boxed{S_i' = \xi_i S_i}$

$\hookrightarrow$ you recover the Curie-Weiss model

$\rightarrow$ Mean-field ferromagnet

Equilibrium phase-diagram



- Retrieval phase: equilibrium confs have a macroscopic overlap with one of the patterns
- Metastable phase: now, eq. confs are not correlated to the patterns !): but if I start with a conf. that is close to a pattern, I will remain close to it.

- Spin-glass: the patterns are indistinguishable from many order stable minimas.

Why is it usefull? patterns carry "physical" information

→ could we infer them rather than the J_{ij}?

Before looking at max-likelihood, let's look to an interesting property.

$$P(\{\vec{s}\}) = \frac{1}{Z} e^{\frac{\beta}{2N}\sum_\mu \left(\sum_i s_i \xi_i^\mu\right)^2}$$

we can "linearize" the sq. in the exponential using the Hubbard-Stratonovitch tf.

$$e^{\frac{a^2x^2}{2}} \propto \int dy\, e^{\underbrace{-\frac{y^2}{2} + ayx}_{= -\frac{1}{2}(y-ax)^2 + \frac{a^2x^2}{2}}}$$

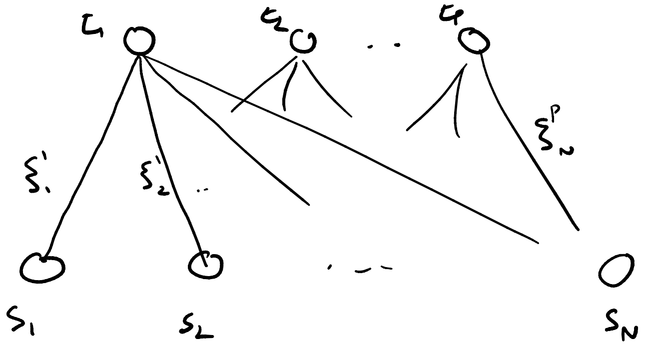
$$p(\{\vec{s}\}) = \frac{1}{Z}\prod_\mu \left[e^{\frac{\beta}{2N}\left(\sum_i \xi_i^\mu s_i\right)^2}\right] = \frac{1}{Z}\prod_\mu \int d\tau_\mu\, e^{-\frac{\tau_\mu^2}{2} + \sqrt{\frac{\beta}{N}}\sum_i \xi_i^\mu s_i \tau_\mu}$$

we can now def the joint distrib. over $\{\vec{s}, \vec{\tau}\}$

$$p(\{\vec{s}\}, \{\vec{\tau}\}) = \frac{1}{Z} \underbrace{\exp\left(-\sum_\mu \frac{\tau_\mu^2}{2}\right)}_{\text{prior dist over } \{\vec{\tau}\}} \cdot \exp\underbrace{\left(\sqrt{\frac{\beta}{N}}\sum_{i,\mu} s_i \xi_i^\mu \tau_\mu\right)}_{-\beta' \mathcal{H}'[\vec{s}, \vec{\tau}]}$$

$$\begin{cases} \beta' = \sqrt{\beta} \\ \mathcal{H}' = -\dfrac{1}{\sqrt{N}} \underbrace{\sum_{i,\mu} s_i \xi_i^\mu \tau_\mu}_{\text{it has a bipartite interacting struct.}} \end{cases}$$

a neural-network topology



here
i) patterns correspond to couplings between $\vec{s}$ and $\vec{\tau}$
ii) $\vec{s}$: visible var
iii) $\vec{\tau}$: hidd var.

In such setting, we want to infer the ξ_i^μ directly.

$$\{\xi_i^\mu\} = \operatorname{argmax}\left[\sum_a \log p(\vec{s}^a)\right] = \operatorname{argmax}\left[\sum_a \underbrace{\left[\overbrace{\int d\vec{\tau}\, e^{-\sum_\mu \frac{\tau_\mu^2}{2} + \sqrt{\frac{\beta}{N}}\sum_{i,\mu} s_i \xi_i^\mu \tau_\mu}}^{\mathcal{L}_d} - \overbrace{\log Z}^{\mathcal{L}_p}\right]}_{\mathcal{L} = \mathcal{L}_d - \mathcal{L}_p}\right]$$

Let's compute the gradient:

$$\frac{\partial \mathcal{L}_d}{\partial \xi_i^\mu} = \underbrace{\frac{1}{D}\sum_a \int d\vec{\tau}\, e^{-\sum_\mu \frac{\tau_\mu^2}{2} + \sqrt{\frac{\beta}{N}}\sum_{i,\mu} s_i^{(a)} \xi_i^\mu \tau_\mu} \cdot s_i^{(a)} \cdot \tau_\mu}_{\frac{1}{D}\sum_a s_i^a \cdot \int d\tau_\mu\, p(\tau_\mu | \vec{s}^{(a)}) \cdot \tau_\mu}$$

$$\text{!!} \qquad = \frac{1}{D}\sum_a s_i^a \left(\sqrt{\frac{\beta}{N}} \sum_j s_j^{(a)} \xi_j^\mu\right) \Leftarrow \langle s_i \tau_\mu \rangle_D$$

$$\frac{\partial \mathcal{L}_p}{\partial \xi_i^\mu} = \langle s_i \tau_\mu \rangle_{p(\vec{s}, \vec{\tau})}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^\mu} = \langle s_i \tau_\mu \rangle_D - \langle s_i \tau_\mu \rangle_{p(\{\vec{s}, \vec{\tau}\})}$$

it depends on the model!

still: we are not matching any high order stat.

Important remark: from the NN reformulation of the Hopfield model, we see that:

1) GMM can be reformulate as a NN.

$$p(x, \vec{z}) = \prod_{h=1}^{c} \left[\rho_h \frac{1}{\sqrt{2\pi\sigma_h^2}} \underbrace{\exp\left[- \frac{(x-\mu_h)^2}{2\sigma_h^2} \right]}\right]^{z_h}$$

In the GMM, the hidden variables follow the softmax distribution

$$\prod_h \exp\left(- \frac{x^2}{2} z_h + \frac{x \mu_h z_h}{\sigma_h^2} \right)$$

$$\shortparallel$$

$$\exp\left(-\frac{x^2}{2}\right) \cdot \exp\left(\sum_h \frac{x \mu_h z_h}{\sigma_h^2} \right)$$

bipartite struct

II) We can generalize the approach by tuning the prior dist over the hidden nodes.

→ using a bernoulli prior we get a $\underbrace{\text{Restricted Boltzmann}}_{\text{RBM}}$ Machine

. a Truncated Gaussian → Relu - RBM