RBM 2hidden

thou

August 2018

1 Introduction

consider a RBM with 2 hidden unit h_1, h_2, the input is σ,which the dimension is N, the inverse temperature is β,so the joint probability is

$$P(\sigma, h_1, h_2) = \frac{1}{Z} \exp \frac{\beta}{\sqrt{N}} (h_1 \sum_{i=1}^{N} \xi_i^1 \sigma_i + h_2 \sum_{i=1}^{N} \xi_i^2 \sigma_i) \tag{1}$$

sum over the hidden unit, we can get,

$$P(\sigma) = \frac{1}{Z} [2 \cosh \frac{\beta}{\sqrt{N}} \sum_{i=1}^{N} (\xi_i^1 + \xi_i^2) \sigma_i + 2 \cosh \frac{\beta}{\sqrt{N}} \sum_{i=1}^{N} (\xi_i^1 - \xi_i^2) \sigma_i] \tag{2}$$

so we can get the partition function as

$$Z = \sum_{\sigma_i} [2 \cosh \frac{\beta}{\sqrt{N}} \sum_{i=1}^{N} (\xi_i^1 + \xi_i^2) \sigma_i + 2 \cosh \frac{\beta}{\sqrt{N}} \sum_{i=1}^{N} (\xi_i^1 - \xi_i^2) \sigma_i] \tag{3}$$

the final results is

$$Z = 2 \prod_{i=1}^{N} [2 \cosh \frac{\beta}{\sqrt{N}} (\xi_i^1 + \xi_i^2)] + 2 \prod_{i=1}^{N} [2 \cosh \frac{\beta}{\sqrt{N}} (\xi_i^1 - \xi_i^2)] \tag{4}$$

Because the systerm scale by $\sqrt{N}$,so we expansion for x,

$$\cosh x = e^{\ln \cosh x} = e^{\ln \frac{1}{2}(e^x + e^{-x})} = e^{\ln (1 + \frac{x^2}{2})} = e^{\frac{x^2}{2}} \tag{5}$$

we simplify the paration function as:

$$Z = 2^{N+1} e^{\beta^2} (2 \cosh \frac{\beta^2}{N} \sum_{i=1}^{N} \xi_i^1 \xi_i^2) \tag{6}$$

the message passing for this systerm is:

$$P_{i\to a}(\xi_i^1,\xi_i^2) \propto \prod_{b\neq a} \hat{P}_{b\to i}(\xi_i^1,\xi_i^2) \tag{7}$$

and

$$\hat{P}_{b\to i}(\xi_i^1,\xi_i^2) \propto \sum_{j\neq i} Z^{-1}\cosh\beta X \cosh\beta Y \prod_{j\neq i} P_{j\to b}(\xi_j^1,\xi_j^2) \tag{8}$$

where

$$X = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \xi_i^1 \sigma_i^a \tag{9}$$

and

$$Y = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \xi_i^2 \sigma_i^a \tag{10}$$

by use

$$\omega_1 = \sum_{j\neq i} \frac{1}{\sqrt{N}} \xi_j^1 \sigma_j^b \tag{11}$$

$$\omega_2 = \sum_{j\neq i} \frac{1}{\sqrt{N}} \xi_j^2 \sigma_j^b \tag{12}$$

by relaxed BP,

$$m_{b\to i}^1 = \sum_{j\neq i} \frac{1}{\sqrt{N}} \sigma_j^n m_{j\to b}^1 \tag{13}$$

$$m_{b\to i}^2 = \sum_{j\neq i} \frac{1}{\sqrt{N}} \sigma_j^n m_{j\to b}^2 \tag{14}$$

and

$$(\hat{\sigma}_{b\to i}^1)^2 = \frac{1}{N}\sum_{j\neq i}[1-(m_{j\to b}^1)^2] \tag{15}$$

$$(\hat{\sigma}_{b\to i}^2)^2 = \frac{1}{N}\sum_{j\neq i}[1-(m_{j\to b}^2)^2] \tag{16}$$

and the correlation is :

$$\rho_{b\to i} = \frac{\frac{1}{N}\sum_{j\in\partial b/i}(Q_{j\to b} - m_{j\to b}^1 m_{j\to b}^2)}{\hat{\sigma}_{b\to i}^1 \hat{\sigma}_{b\to i}^2} \tag{17}$$

So the joint distribution of auxiliary variable ω_1, ω_2 is :

$$P(\omega_1,\omega_2) = e^{-\frac{1}{2}(\omega_1 - m_{b\to i}^1, \omega_2 - m_{b\to i}^2)^T \Sigma^{-1}(\omega_1 - m_{b\to i}^1, \omega_2 - m_{b\to i}^2)}$$

and the precise matrix is :

$$\Sigma^{-1} = \begin{bmatrix} (\hat{\sigma}^1_{b\to i})^2 & \rho_{b\to i}\hat{\sigma}^1_{b\to i}\hat{\sigma}^2_{b\to i} \\ \rho_{b\to i}\hat{\sigma}^1_{b\to i}\hat{\sigma}^2_{b\to i} & (\hat{\sigma}^2_{b\to i})^2 \end{bmatrix}$$

for gaussian integrals, we have ,

$$\int d\boldsymbol{\omega} e^{-\frac{1}{2}\boldsymbol{\omega}^T A\boldsymbol{\omega}+\boldsymbol{\omega}^T\boldsymbol{b}} = \frac{(2\pi)^{\frac{N}{2}}}{\sqrt{\det A}} e^{\frac{1}{2}\boldsymbol{b}^T A^{-1}\boldsymbol{b}} \tag{18}$$

where

$$Q_{j\to b} = \frac{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{+-}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j}}{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{+-}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j}} \tag{19}$$

for factor to node message $\hat{P}_{b\to i}$, consider the terms which include auxiliary variables ω_1, ω_2, we get

$$\int d\boldsymbol{\omega}\cosh[\beta\omega_1+\frac{\beta}{\sqrt{N}}\xi^1_i\sigma^b_i]\cosh[\beta\omega_2+\frac{\beta}{\sqrt{N}}\xi^2_i\sigma^b_i]e^{-\frac{1}{2}(\omega_1-m^1_{b\to i},\omega_2-m^2_{b\to i})^T\Sigma^{-1}(\omega_1-m^1_{b\to i},\omega_2-m^2_{b\to i})}$$

Because $\cosh x\cosh y = \frac{1}{2}(e^{x+y}+e^{x-y}+e^{-x-y}+e^{-x+y})$, consider the symerty , we can calature first term as: by replace the ω_1, ω_1 as $u_1 = \omega_1 - m^1_{b\to i}, u_1 = \omega_1 - m^1_{b\to i}$, we get

$$\int du_1 du_2 e^{\beta u_1+\beta u_2+\beta m^1_{b\to i}+\beta m^2_{b\to i}+\frac{\beta}{\sqrt{N}}\xi^1_i\sigma^b_i+\frac{\beta}{\sqrt{N}}\xi^2_i\sigma^b_i} \times e^{-\frac{1}{2}(u_1,u_2)^T\Sigma^{-1}(u_1,u_2)}$$

using the gaussian integrals

$$\int du_1 du_2 e^{\beta u_1+\beta u_2-\frac{1}{2}(u_1,u_2)^T\Sigma^{-1}(u_1,u_2)} = e^{\frac{1}{2}\beta^2[(\hat{\sigma}^1_{b\to i})^2+\hat{\sigma}^2_{b\to i})^2+2\rho_{b\to i}\sigma^1_{b\to i}\sigma^2_{b\to i}]}$$

we can get

$$\begin{aligned}\hat{P}_{b\to i} = &\cosh^{-1}(\frac{\beta^2}{N}\sum_{j\neq i}Q_{j\to b} + \beta^2\xi^1_i\xi^2_i)e^{\frac{1}{2}\beta^2\{(\hat{\sigma}^1_{b\to i})^2+(\hat{\sigma}^2_{b\to i})^2\}}[e^{\beta^2\rho_{b\to i}\hat{\sigma}^1_{b\to i}\hat{\sigma}^2_{b\to i}}\cosh\beta X \\ &+ e^{-\beta^2\rho_{b\to i}\hat{\sigma}^1_{b\to i}\hat{\sigma}^2_{b\to i}}\cosh\beta Y]\end{aligned} \tag{20}$$

where

$$X = \frac{1}{\sqrt{N}}(\xi^1_i + \xi^2_i)\sigma^b_i + m^1_{b\to i} + m^2_{b\to i} \tag{21}$$

$$Y = \frac{1}{\sqrt{N}}(\xi^1_i - \xi^2_i)\sigma^b_i + m^1_{b\to i} - m^2_{b\to i} \tag{22}$$

and

$$m^1_{j\to b} = \frac{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{+-}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j}}{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j}} \tag{23}$$

$$m^2_{j\to b} = \frac{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j} - \prod_{a\in\partial j/b}\hat{P}^{+-}_{a\to j}}{\prod_{a\in\partial j/b}\hat{P}^{++}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{--}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{+-}_{a\to j} + \prod_{a\in\partial j/b}\hat{P}^{-+}_{a\to j}} \tag{24}$$

These equations have closed for updating.

2 Replica method

We first define some overlaps emergence in the replica equations.

we can define the hamming distance of two true feature as $d = \frac{1-q}{2}$:

$$q = \frac{1}{N}\sum_i \xi_i^{true,1}\xi_i^{true,2}$$

In the calulation , we define:

$$R^{ab}_{11} = \frac{1}{N}\sum_i \xi_i^{a,1}\xi_i^{b,1}$$
$$R^{ab}_{12} = \frac{1}{N}\sum_i \xi_i^{a,1}\xi_i^{b,2}$$
$$R^{ab}_{22} = \frac{1}{N}\sum_i \xi_i^{a,2}\xi_i^{b,2}$$

for each feature set $\boldsymbol{\xi} = (\xi^1, \xi^2)$,we define the sell overlap as :

$$r^{aa} = \frac{1}{N}\sum_i \xi_i^{a,1}\xi_i^{a,2}$$

The inference performance will be measured by the overlap with $\boldsymbol{\xi}^{true}$ as :

$$\begin{aligned} u^a_{11} &= \frac{1}{N}\sum_i \xi_i^{a,1}\xi_i^{true,1} \\ u^a_{22} &= \frac{1}{N}\sum_i \xi_i^{a,2}\xi_i^{true,2} \\ u^a_{12} &= \frac{1}{N}\sum_i \xi_i^{a,1}\xi_i^{true,2} \end{aligned} \tag{25}$$