Perceptron Model with Teacher - Student setting Note

邱俊斌 [1]

(1. 中山大学物理学院物理系, 广州 510000)

2023 年 6 月 5 日

1 模型设定

我们的模型是一个连续感知器模型，神经元的（确定型）输出可以表示为：

$$y_\mu = \mathrm{sign}\left(\frac{1}{\sqrt{N}}\sum_i X_{\mu i} w_i\right) \tag{1}$$

其中，数据 $\mathbf{X} \in \mathbb{R}^{P\times N}$，$N$ 表示数据维度大小，P 是数据量，y 是标签输出。下标 i 表示第 i 个数据维度，μ 表示第 μ 个数据。

在 Teacher-Student 设定下，我们会使用两个网络。第一个网络是固定的，用来生成真实标签，称为 Teacher 网络，网络权重用 $\mathbf{w}^*$ 表示。第二个网络是可学习的，称为 Student 网络。生成一批随机数据 $\mathbf{X} \sim \mathcal{N}(0,1)$，输入 Teacher 网络输出真实标签 y^*。这批数据和真实标签作为训练数据集输入进学生网络进行学习（SGD、AMP 算法等等），训练好后，我们可以利用新的另一批数据作为测试集，从而得到训练误差

$$\epsilon_g = \mathbb{E}_{\mathbf{X}_{\mathbf{new}}\sim\mathcal{N}(0,1)}\left[\delta_{y^*(\mathbf{X}),y(\mathbf{X})}\right] \tag{2}$$

2 理论计算

2.1 贝叶斯最优设定

贝叶斯最优设定实际上包含两个设定：

1. 老师权重的先验等于学生权重的先验 $P(\mathbf{w}^*) = P(\mathbf{w})$
2. 老师网络结构与学生网络的完全一致

在推导 AMP 方程中由于不需要老师网络的信息，所以上面两条设定暂时没用。但是后面 SE 和 Replica 理论推导，以及 Nishimori 恒等式里用到。

我们的目标是求得学生网络中的权重，首先写出学生权重后验概率

$$P(\mathbf{w}|\mathbf{X},\mathbf{y}^*) = \frac{1}{Z_n}\prod_i P_0(w_i)\prod_\mu P_{out}\left(y^*_\mu|\frac{1}{\sqrt{N}}\sum_i X_{\mu i}w_i\right) \tag{3}$$

这时我们可以选择我们想要的估计子用来获得学生权重，例如

$$\hat{\mathbf{w}} = \mathrm{argmax}_{\mathbf{w}}\, P(\mathbf{w}|\mathbf{X},\mathbf{y}^*) \tag{4}$$

2.2 r-BP 算法

整个 BP 算法分为两步，首先 r-BP 算法推导，然后推导 AMP 方程。

首先列出 BP 方程：

$$
\begin{aligned}
m_{i\to\mu}(w_i) =& \frac{1}{Z_{i\to\mu}} P_0(w_i) \prod_{\nu\neq\mu}^{P} \hat{m}_{\nu\to i}(w_i) \\
\hat{m}_{\mu\to i}(w_i) =& \frac{1}{\hat{Z}_{\mu\to i}} \int \prod_{j\neq i}^{N} dw_j P_{out}\left(y_\mu | \frac{1}{\sqrt{N}} \sum_j X_{\mu j} w_j \right) m_{j\to\mu}(w_j)
\end{aligned}
\tag{5}
$$

下一步我们需要对 P_{out} 进行处理。这里有两种处理方式。

2.2.1 定义辅助变量 z_μ

第一种是直接定义辅助量（参考 Lenka2015[2]），$z_\mu = \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i = \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} w_j + \frac{1}{\sqrt{N}} X_{\mu i} w_i$。其中，$\frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} w_j$ 是空腔高斯局域场，根据中心极限定理，其均值和方差记作：

$$
\begin{aligned}
\omega_{\mu\to i} =& \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} \hat{W}_{j\to\mu} \\
V_{\mu\to i} =& \frac{1}{N} \sum_{j\neq i} X_{\mu j}^2 \hat{C}_{j\to\mu}
\end{aligned}
\tag{6}
$$

其中 $\hat{W}_{j\to\mu}$ 和 $\hat{C}_{j\to\mu}$ 是权重 w_j 在空腔概率下的均值和方差：

$$
\begin{aligned}
\hat{W}_{j\to\mu} &= \langle w_j \rangle = \int dw_j m_{j\to\mu}(w_j) w_j \\
\hat{C}_{j\to\mu} &= \langle w_j^2 \rangle - \langle w_j \rangle^2 = \int dw_j m_{j\to\mu}(w_j) w_j^2 - \hat{W}_{j\to\mu}^2
\end{aligned}
\tag{7}
$$

这样子式5化简变成：

$$
\hat{m}_{\mu\to i}(w_i) \propto \int dz_\mu P_{out}(y_\mu | z_\mu) e^{-\frac{\left(z_\mu - \omega_{\mu\to i} - \frac{1}{\sqrt{N}} X_{\mu i} w_i\right)^2}{2V_{\mu\to i}}}
\tag{8}
$$

2.2.2 先对 P_{out} 作傅立叶变换

第二种方法（参考 Lenka2018[1]）先对 P_{out} 作傅立叶变换

$$
P_{out}\left(y_\mu | \frac{1}{\sqrt{n}} \sum_j X_{\mu j} w_j \right) = \frac{1}{2\pi} \int \exp\left[i\xi_\mu \left(\frac{1}{\sqrt{N}} \sum_j X_{\mu j} w_j \right) \hat{P}_{out}(y_\mu, \xi_\mu) \right]
\tag{9}
$$

因此

$$
\hat{m}_{\mu\to i}(w_i) = \frac{1}{2\pi} \frac{1}{\hat{Z}_{\mu\to i}} \int d\xi_\mu \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu i} w_i \right] \hat{P}_{out}(y_\mu, \xi_\mu) \prod_{j\neq i} \int dw_j \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j} w_j \right] m_{j\to\mu}(w_j)
\tag{10}
$$

引入记号 I_j：

$$
I_j = \int dw_j \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j} w_j \right] m_{j\to\mu}(w_j)
\tag{11}
$$

接着对此进行泰勒展开，除去高阶项得:

$$
\begin{aligned}
I_j &= \int dw_j \left(1 + i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j} w_j - \frac{1}{2} i\xi_\mu^2 \frac{1}{N} X_{\mu j}^2 w_j^2\right) m_{j\to\mu}(w_j) \\
&= 1 + i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j} \hat{W}_{j\to\mu} - \frac{1}{2}\xi_\mu^2 \frac{1}{N} X_{\mu j}^2 \hat{C}_{j\to\mu} \\
&= \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j} \hat{W}_{j\to\mu} - \frac{1}{2}\xi_\mu^2 \frac{1}{N} X_{\mu j}^2 \hat{C}_{j\to\mu}\right]
\end{aligned}
\tag{12}
$$

其中

$$
\begin{aligned}
\hat{W}_{j\to\mu} &= \int dw_j m_{j\to\mu}(w_j) w_j \\
\hat{C}_{j\to\mu} &= \int dw_j m_{j\to\mu}(w_j) w_j^2 - \hat{W}_{j\to\mu}^2
\end{aligned}
\tag{13}
$$

这里有处疑问，为何 $\hat{C}$ 的定义要减去 $\hat{W}^2$

整理一下

$$
\hat{m}_{\mu\to i}(w_i) = \frac{1}{2\pi}\frac{1}{\hat{Z}_{\mu\to i}} \int d\xi_\mu \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu i} w_i\right] \hat{P}_{out}(y_\mu, \xi_\mu) \prod_{j\neq i} \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j}\hat{W}_{j\to\mu} - \frac{1}{2}\xi_\mu^2 \frac{1}{N} X_{\mu j}^2 \hat{C}_{j\to\mu}\right]
\tag{14}
$$

最后对 $\hat{P}$ 进行傅立叶变换得到原来的概率分布 P_{out}

$$
\begin{aligned}
\hat{m}_{\mu\to i}(w_i) =& \frac{1}{2\pi}\frac{1}{\hat{Z}_{\mu\to i}} \int dz_\mu P_{out}(y_\mu|z_\mu) \\
& \int d\xi_\mu \exp\left[-i\xi_\mu z_\mu + i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu i} w_i\right] \prod_{j\neq i} \exp\left[i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu j}\hat{W}_{j\to\mu} - \frac{1}{2}\xi_\mu^2 \frac{1}{N} X_{\mu j}^2 \hat{C}_{j\to\mu}\right] \\
=& \frac{1}{2\pi}\frac{1}{\hat{Z}_{\mu\to i}} \int dz_\mu P_{out}(y_\mu|z_\mu) \\
& \int d\xi_\mu \exp\left[-i\xi_\mu z_\mu + i\xi_\mu \frac{1}{\sqrt{N}} X_{\mu i} w_i + i\xi_\mu \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j}\hat{W}_{j\to\mu} - \frac{1}{2}\xi_\mu^2 \frac{1}{N} \sum_{j\neq i} X_{\mu j}^2 \hat{C}_{j\to\mu}\right] \\
=& \frac{1}{2\pi}\frac{1}{\hat{Z}_{\mu\to i}} \int dz_\mu P_{out}(y_\mu|z_\mu) \\
& \int d\xi_\mu \exp\left[-\frac{1}{2}\left(\frac{1}{N}\sum_{j\neq i} X_{\mu j}^2 \hat{C}_{j\to\mu}\right)\xi_\mu^2 + i\left(-z_\mu + \frac{1}{\sqrt{N}} X_{\mu i} w_i + \frac{1}{\sqrt{N}}\sum_{j\neq i} X_{\mu j}\hat{W}_{j\to\mu}\right)\xi_\mu\right] \\
=& \frac{1}{\hat{Z}_{\mu\to i}} \sqrt{\frac{1}{V_{\mu\to i}}} \int dz_\mu P_{out}(y_\mu|z_\mu) e^{-\frac{\left(z_\mu - \omega_{\mu\to i} - \frac{1}{\sqrt{N}} X_{\mu i} w_i\right)^2}{2V_{\mu\to i}}} \\
\propto& \frac{1}{\hat{Z}_{\mu\to i}} \int dz_\mu P_{out}(y_\mu|z_\mu) e^{-\frac{\left(z_\mu - \omega_{\mu\to i} - \frac{1}{\sqrt{N}} X_{\mu i} w_i\right)^2}{2V_{\mu\to i}}}
\end{aligned}
\tag{15}
$$

其中

$$
\begin{aligned}
\omega_{\mu\to i} =& \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} \hat{W}_{j\to\mu} \\
V_{\mu\to i} =& \frac{1}{N} \sum_{j\neq i} X_{\mu j}^2 \hat{C}_{j\to\mu}
\end{aligned}
\tag{16}
$$

可以发现，两个结果式8和式15一样的。

2.2.3 继续计算

接下来需要处理 e 指数，我们发现 $\frac{1}{\sqrt{N}}X_{\mu i}w_i$ 是个小量，可以以此进行展开

$$
\begin{aligned}
H_{\mu\to i} =& \exp\left[-\frac{\left(z_\mu-\omega_{\mu\to i}-\frac{1}{\sqrt{N}}X_{\mu i}w_i\right)^2}{2V_{\mu\to i}}\right] \\
=& \exp\left[-\frac{(z_\mu-\omega_{\mu\to i})^2}{2V_{\mu\to i}}-\frac{-2(z_\mu-\omega_{\mu\to i})\frac{1}{\sqrt{N}}X_{\mu i}w_i+\frac{1}{N}X_{\mu i}^2w_i^2}{2V_{\mu\to i}}\right] \\
=& \exp\left[-\frac{(z_\mu-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right]\left(1+\frac{1}{V_{\mu\to i}}(z_\mu-\omega_{\mu\to i})\frac{1}{\sqrt{N}}X_{\mu i}w_i+\frac{1}{2V_{\mu\to i}^2}(z_\mu-\omega_{\mu\to i})^2\frac{1}{N}X_{\mu i}^2w_i^2\right) \\
& \left(1-\frac{1}{V_{\mu\to i}}\frac{1}{N}X_{\mu i}^2w_i^2\right) \\
=& \exp\left[-\frac{(z_\mu-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right]\left(1+\frac{1}{V_{\mu\to i}}(z_\mu-\omega_{\mu\to i})\frac{1}{\sqrt{N}}X_{\mu i}w_i+\frac{1}{2V_{\mu\to i}^2}(z_\mu-\omega_{\mu\to i})^2\frac{1}{N}X_{\mu i}^2w_i^2\right. \\
& \left.-\frac{1}{V_{\mu\to i}}\frac{1}{N}X_{\mu i}^2w_i^2\right)
\end{aligned}
\tag{17}
$$

汇总起来

$$
\begin{aligned}
\hat{m}_{\mu\to i}(w_i) =& \frac{1}{\hat{Z}_{\mu\to i}}\int dz_\mu P_{out}(y_\mu|z_\mu)\exp\left[-\frac{(z_\mu-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right] \\
& \left(1+\frac{1}{V_{\mu\to i}}(z_\mu-\omega_{\mu\to i})\frac{1}{\sqrt{N}}X_{\mu i}w_i+\frac{1}{2V_{\mu\to i}^2}(z_\mu-\omega_{\mu\to i})^2\frac{1}{N}X_{\mu i}^2w_i^2-\frac{1}{V_{\mu\to i}}\frac{1}{N}X_{\mu i}^2w_i^2\right)
\end{aligned}
\tag{18}
$$

定义测度：

$$
Q_{out}(z;\omega,y,V)=\frac{1}{Z_{out}}\exp\left[-\frac{(z-\omega)^2}{2V}\right]P_{out}(y|z) \tag{19}
$$

其中归一化系数为

$$
Z_{out}(\omega,y,V)=\int dz\exp\left[-\frac{(z-\omega)^2}{2V}\right]P_{out}(y|z) \tag{20}
$$

为了处理上面关于 $(z-\omega)$，$(z-\omega)^2$ 积分，我们记为

$$
g_{out}(\omega,y,V)=\frac{1}{V}\mathbb{E}_{Q_{out}}[z-w] \tag{21}
$$

$$
\partial_\omega g_{out}(\omega,y,V)=\frac{1}{V^2}\mathbb{E}_{Q_{out}}\left[(z-\omega)^2\right]-\frac{1}{V}-g_{out}^2 \tag{22}
$$

因此

$$\begin{aligned}\hat{m}_{\mu\to i}(w_i)=&\frac{1}{\hat{Z}_{\mu\to i}}\Bigg(1+\frac{1}{\sqrt{N}}X_{\mu i}w_i g_{out}(\omega_{\mu\to i},y_\mu,V_{\mu\to i})\\&+\frac{1}{2}\frac{1}{N}X_{\mu i}^2w_i^2\partial_\omega g_{out}(\omega_{\mu\to i},y_\mu,V_{\mu\to i})+\frac{1}{2}\frac{1}{N}X_{\mu i}^2w_i^2g_{out}^2(\omega_{\mu\to i},y_\mu,V_{\mu\to i})\Bigg)\\=&\frac{1}{\hat{Z}_{\mu\to i}}\left(1+B_{\mu\to i}w_i-\frac{1}{2}A_{\mu\to i}w_i^2+\frac{1}{2}B_{\mu\to i}^2w_i^2\right)\\=&\frac{1}{\hat{Z}_{\mu\to i}}\left(1+B_{\mu\to i}w_i-\frac{1}{2}A_{\mu\to i}w_i^2\right)\\=&\frac{1}{\hat{Z}_{\mu\to i}}\exp\left[-\frac{1}{2}A_{\mu\to i}w_i^2+B_{\mu\to i}w_i\right]\end{aligned}\tag{23}$$

其中定义：

$$\begin{aligned}B_{\mu\to i}=&\frac{1}{\sqrt{N}}X_{\mu i}g_{out}(\omega_{\mu\to i},y_\mu,V_{\mu\to i})\\A_{\mu\to i}=&-\frac{1}{N}X_{\mu i}^2\partial_\omega g_{out}(\omega_{\mu\to i},y_\mu,V_{\mu\to i})\end{aligned}\tag{24}$$

这里有个疑问，为何倒数第二舍去 $\frac{1}{2}w_i^2B_{\mu\to i}^2$ 这一项呢

最后可得

$$\begin{aligned}m_{i\to\mu}(w_i)=&\frac{1}{Z_{i\to\mu}}P_0(w_i)\prod_{\nu\neq\mu}\frac{1}{\hat{Z}_{\nu\to i}}\exp\left[-\frac{1}{2}A_{\nu\to i}w_i^2+B_{\nu\to i}w_i\right]\\\propto&\frac{1}{Z_{i\to\mu}}P_0(w_i)\exp\left[-\frac{1}{2}\left(\sum_{\nu\neq\mu}A_{\nu\to i}\right)w_i^2+\left(\sum_{\nu\neq\mu}B_{\nu\to i}\right)w_i\right]\\=&\frac{1}{Z_{i\to\mu}}P_0(w_i)\exp\left[-\frac{1}{2}\frac{1}{\Sigma_{\mu\to i}}w_i^2+\frac{T_{\mu\to i}}{\Sigma_{\mu\to i}}w_i\right]\end{aligned}\tag{25}$$

其中定义

$$\begin{aligned}\Sigma_{\mu\to i}=&\left(\sum_{\nu\neq\mu}A_{\nu\to i}\right)^{-1}\\T_{\mu\to i}=&\frac{\sum_{\nu\neq\mu}B_{\nu\to i}}{\sum_{\nu\neq\mu}A_{\nu\to i}}=\Sigma_{\mu\to i}\left(\sum_{\nu\neq\mu}B_{\nu\to i}\right)\end{aligned}\tag{26}$$

定义测度

$$Q_0(w;\Sigma,T)=\frac{1}{Z_0}P_0(w)\exp\left[-\frac{1}{2}\frac{1}{\Sigma}w^2+\frac{T}{\Sigma}w\right]\propto\frac{1}{Z_0}P_0(w)\exp\left[-\frac{(w-T)^2}{2\Sigma}\right]\tag{27}$$

归一化系数为

$$Z_0(\Sigma,T)=\int dwP_0(w)\exp\left[-\frac{(w-T)^2}{2\Sigma}\right]\tag{28}$$

为了计算上面关于 $\hat{W}_{j\to\mu}$，$\hat{C}_{j\to\mu}$ 积分，我们记为

$$f_w(\Sigma,T)=\mathbb{E}_{Q_0}[w]\tag{29}$$

$$f_c(\Sigma, T) = \mathbb{E}_{Q_0}\left[w^2\right] - f_w^2 \tag{30}$$

$$\begin{aligned}\hat{W}_{i\to\mu} &= \int dw_i m_{i\to\mu}(w_i) w_i = \int dw_i Q_0(w_i; \Sigma_{\mu\to i}, T_{\mu\to i}) w_i = f_w(\Sigma_{\mu\to i}, T_{\mu\to i})\\ \hat{C}_{i\to\mu} &= \int dw_i m_{i\to\mu}(w_i) w_i^2 - \hat{W}_{i\to\mu}^2 = \int dw_i Q_0(w_i; \Sigma_{\mu\to i}, T_{\mu\to i}) w_i^2 - \hat{W}_{i\to\mu}^2 = f_c(\Sigma_{\mu\to i}, T_{\mu\to i})\end{aligned} \tag{31}$$

到此，所有方程自洽闭合。

2.3 AMP 方程

下一步是推导 AMP 方程，这里主要的方式是通过量级分析，加入忽略一些小量，使得每一个物理量都只与一个指标相关，而不需要两个指标，以此减少计算量。

这里我直接列出最后的结果

$$\begin{aligned}\omega_\mu =& \sum_i \left(\frac{1}{\sqrt{N}} X_{\mu i} \hat{W}_i - \frac{1}{N} X_{\mu i}^2 \hat{C}_i g_{out,\mu}\right) = \frac{1}{\sqrt{N}} \sum_i X_{\mu i} \hat{W}_i - V_\mu g_{out,\mu}\\ V_\mu =& \frac{1}{N} \sum_i X_{\mu i}^2 \hat{C}_i\\ g_{out,\mu} =& g_{out}(\omega_\mu, y_\mu, V_\mu)\\ \partial_\omega g_{out,\mu} =& \partial_\omega g_{out}(\omega_\mu, y_\mu, V_\mu)\\ B_\mu =& \frac{1}{\sqrt{N}} X_{\mu i} g_{out,\mu}\\ A_\mu =& -\frac{1}{N} X_{\mu i}^2 \partial_\omega g_{out,\mu}\\ T_i =& \Sigma_i \left(\sum_\mu \left(B_\mu + A_\mu \hat{W}_i\right)\right)\\ \Sigma_i =& \left(\sum_\mu A_\mu\right)^{-1}\\ \hat{W}_i =& f_w(\Sigma_i, T_i)\\ \hat{C}_i =& f_c(\Sigma_i, T_i)\end{aligned} \tag{32}$$

2.4 State Evolution 方程计算

我们会把 AMP 方程中的 $\hat{W}_i$ 视作学生网络中的权重，推导 SE 方程的目标是得到以下定义的序参量

$$
\begin{aligned}
q =& \mathbb{E}_{w^*} \frac{1}{N} \sum_i \left(\hat{W}_i\right)^2 \\
m =& \mathbb{E}_{w^*} \frac{1}{N} \sum_i \hat{W}_i w_i^* \\
Q =& \mathbb{E}_{w^*} \frac{1}{N} \sum_i w_i^{*2} \\
\sigma =& \mathbb{E}_{w^*} \frac{1}{N} \sum_i \hat{C}_i
\end{aligned} \tag{33}
$$

在我们的模型，$Q = 1$ 是个保持不变的量，因此后面推导中会保留 Q 但不会对进行计算。

我们的目标是计算 m，关键需要处理学生网络的权重 $\hat{W}$，由于 $\hat{W}_i = f_w(\Sigma_i, T_i)$，因此下一步是计算 Σ_i 以及 T_i 的统计性质

在此之前先定义一些局域场

$$
\begin{aligned}
\omega_{\mu\to i} =& \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} \hat{W}_{j\to\mu} \\
V_{\mu\to i} =& \frac{1}{N} \sum_{j\neq i} X_{\mu j}^2 \hat{C}_{j\to\mu} \\
z_\mu =& \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^* \\
z_{\mu\to i} =& \frac{1}{\sqrt{N}} \sum_{j\neq i} X_{\mu j} w_j^*
\end{aligned} \tag{34}
$$

其中前两个局域场的定义与 AMP 保持一致，并计算一些后面用到的统计性质

$$
\begin{aligned}
\mathbb{E}_X[\omega_{\mu\to i}\omega_{\mu\to i}] =& \frac{1}{N} \sum_{j,k\neq i} \mathbb{E}_X[X_{\mu j} X_{\mu k}] \hat{W}_{j\to\mu} \hat{W}_{k\to\mu} = \frac{1}{N} \sum_{j\neq i} \hat{W}_{j\to\mu}^2 = q \\
\mathbb{E}_{X,w^*}[z_\mu z_\mu] =& Q \\
\mathbb{E}_{X,w^*}[\omega_{\mu\to i} z_\mu] =& m \\
\mathbb{E}_{X,w^*}[V_{\mu\to i}] =& \sigma
\end{aligned} \tag{35}
$$

接下来有

$$
\begin{aligned}
\frac{T_i}{\Sigma_i} =&\sum_{\mu} B_{\mu\to i} \\
=&\frac{1}{\sqrt{N}}\sum_{\mu} X_{\mu i} g_{out}(\omega_{\mu\to i}, y_\mu, V_{\mu\to i}) \\
=&\frac{1}{\sqrt{N}}\sum_{\mu} X_{\mu i} g_{out}\left(\omega_{\mu\to i}, \phi_{out}(\frac{1}{\sqrt{N}}\sum_{j\neq i} X_{\mu j} w_j^* + \frac{1}{\sqrt{N}} X_{\mu i} w_i^*), V_{\mu\to i}\right) \\
=&\frac{1}{\sqrt{N}}\sum_{\mu} X_{\mu i}\left(g_{out}\left(\omega_{\mu\to i}, \phi_{out}(\frac{1}{\sqrt{N}}\sum_{j\neq i} X_{\mu j} w_j^*), V_{\mu\to i}\right) + \partial_z g_{out}(\omega_{\mu\to i}, \phi_{out}(z), V_{\mu\to i})\frac{1}{\sqrt{N}} X_{\mu i} w_i^*\right) \\
=&\frac{1}{\sqrt{N}}\sum_{\mu} X_{\mu i} g_{out}\left(\omega_{\mu\to i}, \phi_{out}(\frac{1}{\sqrt{N}}\sum_{j\neq i} X_{\mu j} w_j^*), V_{\mu\to i}\right) + \frac{1}{N}\sum_{\mu} X_{\mu i}^2 w_i^* \partial_z g_{out}(\omega_{\mu\to i}, \phi_{out}(z), V_{\mu\to i})
\end{aligned}
\tag{36}
$$

定义

$$
\begin{aligned}
\hat{q} =&\mathbb{E}_{\omega,z}\left[g_{out}^2(\omega, \phi_{out}(z), V)\right] \\
\hat{m} =&\mathbb{E}_{\omega,z}[\partial_z g_{out}(\omega, \phi_{out}(z), V)] \\
\hat{\chi} =&\mathbb{E}_{\omega,z}[-\partial_\omega g_{out}(\omega, \phi_{out}(z), V)]
\end{aligned}
\tag{37}
$$

而 $\frac{T_i}{\Sigma}$ 中的第一项均值 0，方差 $\alpha\hat{q}$，方差主导，可以重参数化技巧，用一个高斯变量表示，第二项的均值 $\alpha\hat{m}w_i^*$，可以直接使用均值表示。因此

$$
\mathbb{E}_{\omega,z}\left[\frac{T_i}{\Sigma_i}\right] =\sqrt{\alpha\hat{q}}\xi + \alpha\hat{m}w_i^* \tag{38}
$$

其中 ξ 一个 $\mathcal{N}(0,1)$ 随机变量。

除此以外

$$
\Sigma_i^{-1} = -\frac{1}{N}\sum_{\mu} X_{\mu i}^2 \partial_\omega g_{out}(\omega_\mu, y_\mu, V_\mu) = \alpha\hat{\chi} \tag{39}
$$

因此

$$
\begin{aligned}
q =&\mathbb{E}_{w^*,\Sigma,T}\left[f_w^2(\Sigma, T)\right] \\
=&\int dw^* P_0(w^*)\int D\xi f_w^2(\frac{1}{\alpha\chi}, \sqrt{\frac{\hat{q}}{\alpha\hat{\chi}^2}}\xi + \frac{\hat{m}}{\hat{\chi}})
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
\hat{\chi} =& -\mathbb{E}_{w^*,\omega,z,V}[\partial_\omega g_{out}(\omega, \phi_{out}(z), V)] \\
=& -\int d\omega \frac{e^{-\frac{1}{2}\frac{\omega^2}{q}}}{\sqrt{2\pi q}} dz \frac{e^{-\frac{1}{2}\frac{(z-\omega)^2}{Q-q}}}{\sqrt{2\pi(Q-q)}}\partial_\omega g_{out}(\omega, \phi_{out}(z), \sigma)
\end{aligned}
\tag{41}
$$

2.4.1 Nishimori 恒等式子

在贝叶斯最优的情况下，可以证明 Nishimori 恒等式子，部分序参量有相等的性质，即是

$$
\begin{aligned}
q =&m \\
\hat{q} =&\hat{m} = \hat{\chi} \\
\sigma =&Q - q
\end{aligned}
\tag{42}
$$

以此化简两个 SE 方程

$$
\begin{aligned}
m &= \int dw^* P_0(w^*) \int D\xi f_w^2(\frac{1}{\alpha m}, \frac{\xi}{\alpha \hat{m}} + w^*) \\
\hat{m} &= -\int d\omega \frac{e^{-\frac{1}{2}\frac{\omega^2}{m}}}{\sqrt{2\pi m}} dz \frac{e^{-\frac{1}{2}\frac{(z-\omega)^2}{Q-m}}}{\sqrt{2\pi(Q-m)}} \partial_\omega g_{out}(\omega, \phi_{out}(z), Q-m)
\end{aligned} \tag{43}
$$

2.4.2 泛化误差计算

泛化误差可以有许多定义，如果以二分类泛化误差（模型设定保持一致）

$$
\epsilon_g = \mathbb{E}_{\mathbf{X}^{\mathbf{new}} \sim \mathcal{N}(0,1)} \left[\text{sign}\left(\frac{1}{\sqrt{N}} \sum_i X_{\mu i}^{new} \hat{W}_i \right) == \text{sign}\left(\frac{1}{\sqrt{N}} \sum_i X_{\mu i}^{new} w_i^* \right) \right] \tag{44}
$$

对上式进行重参数化可以得

$$
\epsilon_g = \int DxDyDz \Big[\text{sign}\left(\sqrt{m}x + \sqrt{q-m}y\right) == \text{sign}\left(\sqrt{m}x + \sqrt{Q-m}z\right) \Big] \tag{45}
$$

考虑 Nishimori 条件，可得

$$
\epsilon_g = \int DxDy \Big[\text{sign}\left(\sqrt{m}x\right) == \text{sign}\left(\sqrt{m}x + \sqrt{Q-m}y\right) \Big] \tag{46}
$$

其中 $[a == b]$ 等价 $\delta_{a,b}$

2.5 Replica 计算

根据权重后验概率公式，我们可以写出配分函数

$$
Z = \int \prod_\mu dy_\mu \prod_i dw_i P_0(w_i) \prod_\mu P_{out}\left(y_\mu | \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i \right) \tag{47}
$$

$$
Z^n = \int \prod_\mu dy_\mu \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}\left(y_\mu | \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^a \right) \tag{48}
$$

为了方便表示老师权重，我们使用指标 0 表示老师网络，并用 Z^{n+1} 表示加入老师网络后的配分函数，即：

$$
Z^{n+1} = \int \prod_\mu dy_\mu \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}\left(y_\mu | \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^a \right) \tag{49}
$$

注意后面在鞍点近似操作还是除 n，从而 $n \to 0$ 时会出现 $\frac{1}{n} \ln \int I^{n+1} \to \int I \ln I$

定义辅助场 $z_\mu^a = \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^a$，而且有：

$$
\left\langle z_\mu^a \right\rangle = 0, \quad \left\langle z_\mu^a z_\nu^b \right\rangle = \delta_{\mu\nu} \frac{1}{N} \sum_i w_i^a w_i^b = Q^{ab} \tag{50}
$$

$$
\begin{aligned}
Z^{n+1} &= \int \prod_\mu dy_\mu \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}(y_\mu | z_\mu^a) \int \left(\prod_{\mu a} dz_\mu^a \right) \prod_{\mu a} \delta\left(z_\mu^a - \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^a \right) \\
&= \int \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}(y_\mu | z_\mu^a) \int \left(\prod_{\mu a} \frac{dz_\mu^a d\hat{z}_\mu^a}{2\pi} \right) \prod_{\mu a} e^{-i z_\mu^a \hat{z}_\mu^a + i \hat{z}_\mu^a \frac{1}{\sqrt{N}} \sum_i X_{\mu i} w_i^a}
\end{aligned} \tag{51}
$$

$$\left\langle Z^{n+1}\right\rangle = \prod_{\mu} dy_{\mu} \int \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) \int \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \left\langle \prod_{\mu i} e^{i\frac{1}{\sqrt{N}}\sum_a \hat{z}_{\mu}^a X_{\mu i} w_i^a} \right\rangle \tag{52}$$

其中平均项为:

$$\left\langle e^{i\hat{z}_{\mu}^a \frac{1}{\sqrt{N}}\sum_i X_{\mu i} w_i^a} \right\rangle = e^{-\frac{1}{2}\frac{1}{N}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b w_i^a w_i^b} \tag{53}$$

因此可得

$$\begin{aligned}
\left\langle Z^{n+1}\right\rangle =& \int \prod_{\mu} dy_{\mu} \prod_{ia} dw_i^a P_0(w_i^a) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) \int \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \prod_{\mu i} e^{-\frac{1}{2}\frac{1}{N}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b w_i^a w_i^b} \\
=& \int \prod_{\mu} dy_{\mu} \prod_{ia} dw_i^a P_0(w_i^a) \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \prod_{\mu} e^{-\frac{1}{2}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b \frac{1}{N}\sum_i w_i^a w_i^b} \\
=& \int \prod_{\mu} dy_{\mu} \prod_{ia} dw_i^a P_0(w_i^a) \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \prod_{\mu} e^{-\frac{1}{2}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b Q^{ab}} \\
& \int \prod_{ab} dQ^{ab} \delta\left(Q^{ab} - \frac{1}{N}\sum_i w_i^a w_i^b \right) \\
=& \int \prod_{\mu} dy_{\mu} \prod_{ia} dw_i^a P_0(w_i^a) \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \prod_{\mu} e^{-\frac{1}{2}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b Q^{ab}} \\
& \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-i\sum_{ab} Q^{ab}\hat{Q}^{ab} + i\sum_{ab}\hat{Q}^{ab}\frac{1}{N}\sum_i w_i^a w_i^b} \\
=& \int \prod_{\mu} dy_{\mu} \prod_{ia} dw_i^a P_0(w_i^a) \left(\prod_{\mu a} \frac{dz_{\mu}^a d\hat{z}_{\mu}^a}{2\pi} \right) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) e^{-i\sum_{\mu a} z_{\mu}^a \hat{z}_{\mu}^a} \prod_{\mu} e^{-\frac{1}{2}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b Q^{ab}} \\
& \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-N\sum_{ab} Q^{ab}\hat{Q}^{ab} + \sum_{ab}\hat{Q}^{ab}\sum_i w_i^a w_i^b}
\end{aligned} \tag{54}$$

接下来可以积分去 $\hat{z}$

$$\int \prod_{\mu a} d\hat{z}_{\mu}^a \prod_{\mu} e^{-i\sum_a z_{\mu}^a \hat{z}_{\mu}^a - \frac{1}{2}\sum_{ab}\hat{z}_{\mu}^a \hat{z}_{\mu}^b Q^{ab}} = \prod_{\mu} \frac{1}{\sqrt{(2\pi)^{n+1}\det Q}} e^{-\frac{1}{2}\sum_{ab} z_{\mu}^a \tilde{Q}^{ab} z_{\mu}^b} \tag{55}$$

其中记号 $\tilde{Q} = Q^{-1}$

汇总一下

$$\begin{aligned}
\left\langle Z^{n+1}\right\rangle =& \prod_{\mu} dy_{\mu} \int \prod_{ia} dw_i^a P_0(w_i^a) \left(\prod_{\mu a} \frac{dz_{\mu}^a}{2\pi} \right) \prod_{\mu a} P_{out}\big(y_{\mu}|z_{\mu}^a\big) \prod_{\mu} \frac{1}{\sqrt{(2\pi)^{n+1}\det Q}} e^{-\frac{1}{2}\sum_{ab} z_{\mu}^a \tilde{Q}^{ab} z_{\mu}^b} \\
& \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-N\sum_{ab} Q^{ab}\hat{Q}^{ab} + \sum_{ab}\hat{Q}^{ab}\sum_i w_i^a w_i^b} \\
\propto& \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-N\sum_{ab} Q^{ab}\hat{Q}^{ab} + \sum_i \ln\int\prod_a dw_i^a P_0(w_i^a) e^{\sum_{ab}\hat{Q}^{ab}\sum_i w_i^a w_i^b}} \\
& e^{\sum_{\mu}\ln\int\prod_{\mu} dy_{\mu}\prod_a dz_{\mu}^a \prod_a P_{out}\left(y_{\mu}|z_{\mu}^a\right) e^{-\frac{1}{2}\sum_{ab} z_{\mu}^a \tilde{Q}^{ab} z_{\mu}^b - \frac{1}{2}\sum_{\mu}\ln\det Q}} \\
=& \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-N\left(\sum_{ab} Q^{ab}\hat{Q}^{ab} + I + \alpha J\right)}
\end{aligned} \tag{56}$$

其中记为

$$
\begin{aligned}
I &= \ln \int \prod_a dw^a P_0(w^a) e^{\sum_{ab} \hat{Q}^{ab} w^a w^b} \\
J &= \ln \int dy \prod_a dz^a \prod_a P_{out}(y|z^a) e^{-\frac{1}{2}\sum_{ab} z^a \tilde{Q}^{ab} z^b - \frac{1}{2}\ln\det Q}
\end{aligned} \tag{57}
$$

这里略去部分常数。

RS 对称假设下有，$Q^{ab} = Q\delta^{ab} + q(1-\delta ab)$，$\hat{Q}^{ab} = \hat{Q}\delta^{ab} + \hat{q}\big(1-\delta^{ab}\big)$，此时有些特殊结果

$$
\det Q = (Q-q)^{n-1}(Q+(n-1)q) \to 1 \tag{58}
$$

$$
\tilde{Q}^{aa} = \frac{Q+(n-2)q}{(Q-q)(Q+(n-1)q)} \to \frac{Q-2q}{(Q-q)^2} \tag{59}
$$

$$
\tilde{Q}^{ab} = \frac{-q}{(Q-q)(Q+(n-1)q)} \to -\frac{q}{(Q-q)^2}, a \neq b \tag{60}
$$

因此

$$
\langle Z^{n+1} \rangle = \int \prod_{ab} dQ^{ab} d\hat{Q}^{ab} e^{-N\left((P+1)Q\hat{Q} + P(P+1)q\hat{q} + I + \alpha J\right)} \tag{61}
$$

$$
\begin{aligned}
I &= \ln \int \prod_a dw^a P_0(w^a) e^{\sum_{ab} \hat{Q}^{ab} w^a w^b} \\
&= \ln \int \prod_a dw^a P_0(w^a) e^{\frac{\hat{q}}{2}\left(\sum_a w^a\right)^2 - \frac{1}{2}(\hat{Q}-\hat{q})\sum_a (w^a)^2} \\
&= \ln \int D\xi \prod_a dw^a P_0(w^a) e^{\sqrt{\hat{q}}\sum_a w^a \xi - \frac{1}{2}(\hat{Q}-\hat{q})\sum_a (w^a)^2} \\
&= \ln \int D\xi \Big[\int dw P_0(w) e^{\sqrt{\hat{q}} w\xi - \frac{1}{2}(\hat{Q}-\hat{q})w^2}\Big]^{P+1}
\end{aligned} \tag{62}
$$

$$
\begin{aligned}
J &= \ln \int dy \prod_a dz^a \prod_a P_{out}(y|z^a) e^{-\frac{1}{2}\sum_{ab} z^a \tilde{Q}^{ab} z^b} \\
&= \ln \int dy \prod_a dz^a \prod_a P_{out}(y|z^a) e^{-\tilde{q}\left(\sum_a z^a\right)^2 - \frac{1}{2}\left(\tilde{Q}-\tilde{q}\right)\sum_a (z^a)^2} \\
&= \ln \int D\xi dy \prod_a dz^a \prod_a P_{out}(y|z^a) e^{i\sqrt{\tilde{q}}\sum_a z^a \xi - \frac{1}{2}\left(\tilde{Q}-\tilde{q}\right)\sum_a (z^a)^2} \\
&= \ln \int D\xi dy \Big[\int dz P_{out}(y|z) e^{i\sqrt{\tilde{q}} z\xi - \frac{1}{2}\left(\tilde{Q}-\tilde{q}\right)z^2}\Big]^{P+1}
\end{aligned} \tag{63}
$$

这里有部分计算过程与参考文献不太一样，但结果是一样

最终可得

$$
-\beta f = -\frac{1}{2} q\hat{q} + I + \alpha J \tag{64}
$$

$$
I = \ln \int D\xi \int dw^* P_0(w^*) e^{\sqrt{\hat{q}} w^* \xi - \frac{1}{2}\hat{q} w^{*2}} \ln \int dw P_0(w) e^{\sqrt{\hat{q}} w\xi - \frac{1}{2}\hat{q} w^2} \tag{65}
$$

$$
\begin{aligned}
J &= \ln \int D\xi dy \int dz P_{out}(y|z) e^{i\sqrt{\tilde{q}} z\xi - \frac{1}{2}\left(\tilde{Q}-\tilde{q}\right)z^2} \ln \int dz P_{out}(y|z) e^{i\sqrt{\tilde{q}} z\xi - \frac{1}{2}\left(\tilde{Q}-\tilde{q}\right)z^2} \\
&= \ln \int D\xi dy \int Dz P_{out}\Big(y|\sqrt{Q-q}z + \sqrt{q}\xi\Big) \ln \int Dz P_{out}\Big(y|\sqrt{Q-q}z + \sqrt{q}\xi\Big)
\end{aligned} \tag{66}
$$

鞍点方程计算放在后面给出结果

3 感知机下对方程的计算

在感知机模型中，

$$P_0(w) = \frac{e^{-\frac{1}{2}w^2}}{\sqrt{2\pi}} \tag{67}$$

$$P_{out}(y|z) = \delta(y - \mathrm{sign}\,(z)) \tag{68}$$

$$\phi_{out}(z) = \mathrm{sign}\,(z) \tag{69}$$

3.1 AMP 方程

带入上面部分方程可以化简计算

$$Z_{out}(\omega, y, V) = \sqrt{\frac{\pi}{2}}\sqrt{V}\left(1 + y * \mathrm{erf}\left(\frac{\omega}{\sqrt{2V}}\right)\right) \tag{70}$$

$$g_{out}(\omega, y, V) = \frac{y * e^{\frac{\omega^2}{2V}}}{Z_{out}(\omega, y, V)} \tag{71}$$

$$\partial_\omega g_{out}(\omega, y, V) = \frac{\sqrt{\frac{\pi}{2}}V^{\frac{3}{2}}(1 + y * \mathrm{erf}\left(\frac{\omega}{\sqrt{2V}}\right)) + y * e^{-\frac{\omega^2}{2V}}V\omega}{V^2 Z_{out}(\omega, y, V)} \tag{72}$$

$$f_w(\Sigma, T) = \frac{T}{1+\Sigma} \tag{73}$$

$$f_c(\Sigma, T) = \frac{\Sigma}{1+\Sigma} \tag{74}$$

3.2 Replica 计算

$$I = \ln \int D\xi \frac{e^{\frac{\hat{q}\xi^2}{2(1+\hat{q})}}}{\sqrt{1+\hat{q}}} \ln \frac{e^{\frac{\hat{q}\xi^2}{2(1+\hat{q})}}}{\sqrt{1+\hat{q}}} = \ln \int D\xi \hat{t} \ln \hat{t} \tag{75}$$

其中记号 $\hat{t} = \frac{e^{\frac{\hat{q}\xi^2}{2(1+\hat{q})}}}{\sqrt{1+\hat{q}}}$

注意感知机模型 $\int dy \to \sum_{y=\pm 1}$，有

$$\begin{aligned} J = &\ln \int D\xi \left(\frac{1}{2}(1 + \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)) \ln \frac{1}{2}\left(1 + \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)\right)\right. \\ &\left. + \frac{1}{2}\left(1 - \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)\right) \ln \frac{1}{2}\left(1 - \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)\right)\right) \\ = &\ln \int D\xi (t_+ \ln t_+ + t_- \ln t_-) \end{aligned} \tag{76}$$

其中记号 $t_+ = \frac{1}{2}\left(1 + \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)\right)$，$t_- = \frac{1}{2}\left(1 - \mathrm{erf}\left(\sqrt{\frac{q}{2(Q-q)}}\xi\right)\right)$

鞍点方程

$$q = 2\int D\xi e^{\hat{t}} \frac{e^{\hat{t}}(1+\hat{q}-\xi^2)\left(1+\ln \hat{t}\right)}{2(1+\hat{q})^{\frac{5}{2}}} \tag{77}$$

$$\hat{q} = 2\alpha \int D\xi \frac{e^{-\frac{1}{2}\frac{q}{Q-q}\xi^2} Q\xi(\ln t_+ - \ln t_-)}{2\sqrt{2\pi}(Q-q)^2\sqrt{\frac{q}{Q-q}}} \tag{78}$$

4 实验结果

这次同时三种方法进行模拟，实验结果

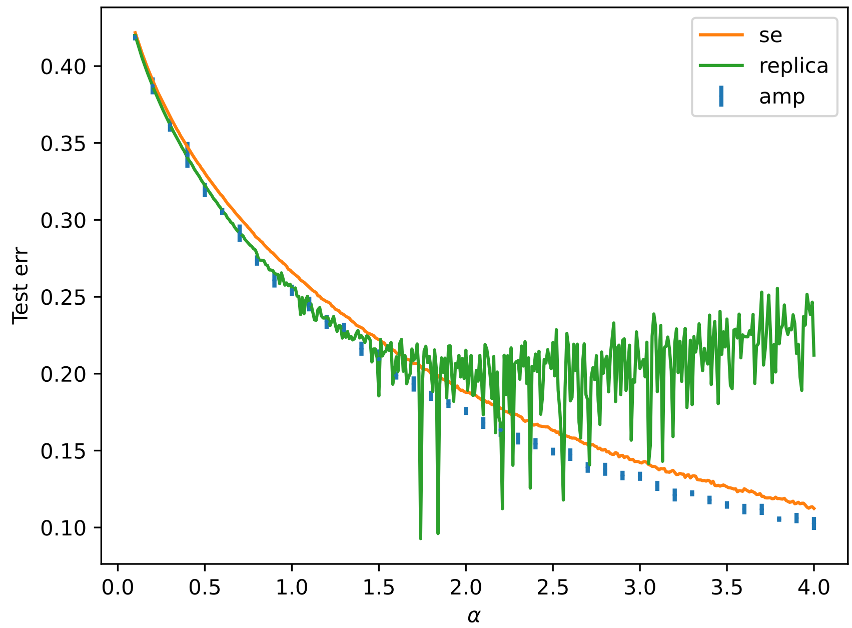

图 1:

目前实验有个问题，replica 模拟部分 α 较大时迭代失败，可能是 MC 积分计算，对于函数奇点处理不好导致积分不准

参考文献

[1] Benjamin Aubin, Antoine Maillard, Jean Barbier, Florent Krzakala, Nicolas Macris, and Lenka Zdeborová. The committee machine: computational to statistical gaps in learning a two-layers neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2019:124023, 12 2019.

[2] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: thresholds and algorithms. *Advances in Physics*, 65:453–552, 9 2016.