连续权重感知机在二分类任务中的泛化误差

LYH & WYC
PMI Group Meeting
June, 17, 2023

Introduction

teacher

$$y = \mathrm{sign}\left(\frac{1}{\sqrt{n}}\mathbf{X}\mathbf{w}^\star\right)$$

$$P_{\mathbf{w}^*} \sim \mathcal{N}(0,1)$$

student

$$P(\mathbf{w}|\mathbf{y},\mathbf{X}) = \frac{P(\mathbf{y}|\mathbf{X},\mathbf{w})P(\mathbf{w})}{P(\mathbf{y},\mathbf{X})} = \frac{1}{\mathcal{Z}}P(\mathbf{y}|\mathbf{X},\mathbf{w})P(\mathbf{w})$$

$$P(\mathbf{y}|\mathbf{X},\mathbf{w}) = \mathbb{1}\left[y = \mathrm{sign}\left(\frac{1}{\sqrt{n}}\mathbf{X}\mathbf{w}\right)\right]$$

generalization error

$$\varepsilon_{\mathrm{gen}} = \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\Theta(-\mathbf{y}\hat{\mathbf{y}})\right] = \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\Theta(-\mathbf{z}\hat{\mathbf{z}})\right]$$

$$= \frac{1}{\pi}\arccos\left(\frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}}{\sqrt{\rho_{\mathbf{w}^\star}\sigma_{\hat{\mathbf{w}}}}}\right)$$

$$= \frac{1}{\pi}\arccos\left(\sqrt{\frac{q}{\rho_{\mathbf{w}^\star}}}\right)$$

Nishimori condition

$$\sigma_{\mathbf{w}^\star\hat{\mathbf{w}}} = \lim_{n\to\infty}\mathbb{E}_{\mathbf{w}^\star,\mathrm{X}}\frac{1}{n}\hat{\mathbf{w}}^\top\mathbf{w}^\star = \sigma_{\hat{\mathbf{w}}} = \lim_{n\to\infty}\mathbb{E}_{\mathbf{w}^\star,\mathrm{X}}\frac{1}{n}\|\hat{\mathbf{w}}\|_2^2$$

$$\rho_{\mathbf{w}^\star} = \sigma_{\mathbf{w}^*} \equiv \lim_{n\to\infty}\mathbb{E}_{\mathbf{w}^\star}\frac{1}{n}\|\mathbf{w}^\star\|_2^2$$

$$\alpha = d/n$$

$$\mathbf{z} = \mathbf{X}\mathbf{w}^\star/\sqrt{n}, \quad \hat{\mathbf{z}} = \mathbf{X}\hat{\mathbf{w}}/\sqrt{n}$$

$$\sigma = \lim_{n\to\infty}\mathbb{E}_{\mathbf{w}^*,X}\frac{1}{n}\begin{bmatrix}\mathbf{w}^{*\top}\mathbf{w}^* & \mathbf{w}^{*\top}\hat{\mathbf{w}}\\ \mathbf{w}^{*\top}\hat{\mathbf{w}} & \hat{\mathbf{w}}^\top\hat{\mathbf{w}}\end{bmatrix} \equiv \begin{bmatrix}\sigma_{\mathbf{w}^*} & \sigma_{\mathbf{w}^*\hat{\mathbf{w}}}\\ \sigma_{\mathbf{w}^*\hat{\mathbf{w}}} & \sigma_{\hat{\mathbf{w}}}\end{bmatrix}$$

$$\hat{\mathbf{z}} = \sqrt{\sigma_{\hat{\mathbf{w}}}}\mathbf{x}_1$$

$$\mathbf{z} = \sqrt{\frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_1 + \sqrt{\sigma_{\mathbf{w}^*}^2 - \frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_2$$

$$\varepsilon_{\mathrm{gen}} = \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\Theta(-\mathbf{z}\hat{\mathbf{z}})\right]$$

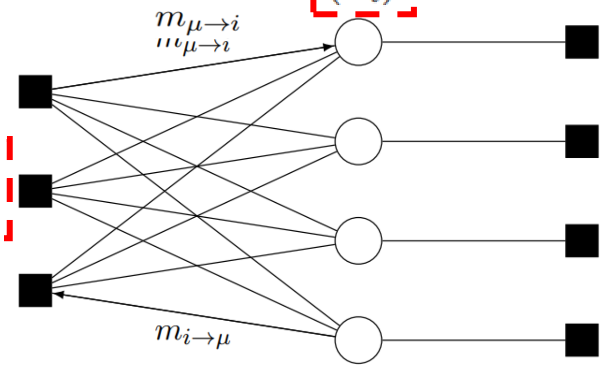
$$= \int D\mathbf{x}_1\int D\mathbf{x}_2\Theta\left(-\sqrt{\sigma_{\hat{\mathbf{w}}}}\mathbf{x}_1\left(\sqrt{\frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}^2}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_1 + \sqrt{\sigma_{\mathbf{w}^*}^2 - \frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}^2}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_2\right)\right)$$

$$= \int D\mathbf{x}_1\int D\mathbf{x}_2\Theta(-\mathbf{x}_1)\,\Theta\left(\sqrt{\frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}^2}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_1 + \sqrt{\sigma_{\mathbf{w}^*}^2 - \frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}^2}{\sigma_{\hat{\mathbf{w}}}}}\mathbf{x}_2\right)$$

$$= \frac{1}{\pi}\arccos\left(\frac{\sigma_{\mathbf{w}^*\hat{\mathbf{w}}}}{\sqrt{\rho_{\mathbf{w}^\star}\sigma_{\hat{\mathbf{w}}}}}\right)$$

AMP

(w_i)



$P_{\text{out}}\left(y_\mu \mid \frac{1}{\sqrt{n}}\sum_{l=1}^{n} w_l X_{l\mu}\right)$

$P_0\left(w_i\right)$

$$\mathcal{H}(\boldsymbol{x};\boldsymbol{Y},\boldsymbol{\Phi}) := -\sum_{\mu=1}^{M} \ln P_{\text{out}}\left(Y_\mu \mid \frac{1}{\sqrt{N}}[\boldsymbol{\Phi}\boldsymbol{x}]_\mu\right)$$

$$E(\boldsymbol{x}) = -\sum_a \ln f_a(\boldsymbol{x}_a)$$

$$P_{a\to i}(x_i) = \sum_{\boldsymbol{x}_j : j\in\partial a\backslash i} f_a(\boldsymbol{x}_a) \prod_{j\in\partial a\backslash i} P_{j\to a}(x_j)$$

$$P_{i\to a}(x_i) = \frac{1}{Z_{i\to a}} \prod_{b\in\partial i\backslash a} P_{b\to i}(x_i).$$

BP equation

$$m_{i\to\mu}\left(w_i\right) = \frac{1}{z_{i\to\mu}} P_0\left(w_i\right) \prod_{\gamma\neq\mu} m_{\gamma\to i}\left(w_i\right)$$

$$m_{\mu\to i}\left(w_i\right) = \frac{1}{z_{\mu\to i}} \int \prod_{j\neq i} \mathrm{d}w_j P_{\text{out}}\left(y_\mu \mid \frac{1}{\sqrt{n}}\sum_{l=1}^{n} w_l X_{l\mu}\right) m_{j\to\mu}\left(w_j\right)$$

relaxed-BP equation

approximate $m_{i\to\mu}$ only TWO moments

Central Limit Theory

$$z_\mu = X_{\mu i} w_i + \sum_{j\neq i} X_{\mu j} w_j$$

$$\sum_{j\neq i} X_{\mu j} w_j \sim \mathcal{N}\left(\omega_{\mu\to i}, V_{\mu\to i}\right)$$

$$\omega_{\mu\to i} = \sum_{j\neq i} X_{\mu j} \hat{w}_{j\to\mu}$$

$$V_{\mu\to i} = \sum_{j\neq i} X_{\mu j}^2 v_{j\to\mu}$$

$$\hat{w}_{i\to\mu} \equiv \int \mathrm{d}w_i\, m_{i\to\mu}\left(w_i\right) w_i$$

$$v_{i\to\mu} \equiv \int \mathrm{d}w_i\, m_{i\to\mu}\left(x_i\right) w_i^2 - \hat{w}_{i\to\mu}^2$$

$$m_{\mu\to i}\left(w_i\right) \propto \int dz_\mu P_{\text{out}}\left(y_\mu \mid z_\mu\right) \exp\left\{-\frac{\left(z_\mu - \omega_{\mu\to i} - X_{\mu i} w_i\right)^2}{2V_{\mu\to i}}\right\}$$

AMP

$$m_{\mu\to i}(w_i) \propto \int dz_\mu P_{\text{out}}\left(y_\mu|z_\mu\right)\exp\left\{-\frac{(z-\omega_{\mu\to i}-X_{\mu i}w_i)^2}{2V_{\mu\to i}}\right\}$$

$$\begin{aligned}
&\exp\left\{-\frac{(z-\omega_{\mu\to i}-X_{\mu i}w_i)^2}{2V_{\mu\to i}}\right\}\\
&=\exp\left\{-\frac{(z-\omega_{\mu\to i})^2+X_{\mu i}^2w_i^2-2\left(z-\omega_{\mu\to i}\right)X_{\mu i}w_i}{2V_{\mu\to i}}\right\}\\
&=\exp\left\{-\frac{(z-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right\}\exp\left\{-\frac{X_{\mu i}^2w_i^2-2\left(z-\omega_{\mu\to i}\right)X_{\mu i}w_i}{2V_{\mu\to i}}\right\}\\
&=\exp\left\{-\frac{(z-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right\}\left(1+X_{\mu i}^2w_i^2-2\left(z-\omega_{\mu\to i}\right)X_{\mu i}w_i+\frac{1}{2}\left(z-\omega_{\mu\to i}\right)^2X_{\mu i}^2w_i^2+\mathcal{O}(\frac{1}{n})\right)
\end{aligned}$$

$$m_{\mu\to i}(w_i) \propto \int dz_\mu P_{\text{out}}\left(y_\mu|z_\mu\right)\exp\left\{-\frac{(z-\omega_{\mu\to i})^2}{2V_{\mu\to i}}\right\}\times\left(1+X_{\mu i}^2w_i^2-2(z-\omega_{\mu\to i})X_{\mu i}w_i+\frac{1}{2}(z-\omega_{\mu\to i})^2X_{\mu i}^2w_i^2+\mathcal{O}(\frac{1}{n})\right)$$

define

$$f_{\text{out}}(\omega,y,V)\equiv\frac{\int \mathrm{d}zP_{\text{out}}\left(y|z\right)(z-\omega)e^{-\frac{(z-\omega)^2}{2V}}}{V\int \mathrm{d}zP_{\text{out}}\left(y|z\right)e^{-\frac{(z-\omega)^2}{2V}}}$$

$$\partial_\omega f_{\text{out}}(\omega,y,V)=\frac{\int \mathrm{d}zP_{\text{out}}\left(y|z\right)(z-\omega)^2e^{-\frac{(z-\omega)^2}{2V}}}{V^2\int \mathrm{d}zP_{\text{out}}\left(y|z\right)e^{-\frac{(z-\omega)^2}{2V}}}-\frac{1}{V}-f_{\text{out}}^2(\omega,y,V)$$

$$A_{\mu\to i}^t=-X_{\mu i}^2\,\partial_\omega f_{\text{out}}\left(\omega_{\mu\to i}^t,y_\mu,V_{\mu\to i}^t\right)$$

$$B_{\mu\to i}^t=X_{\mu i}\,f_{\text{out}}\left(\omega_{\mu\to i}^t,y_\mu,V_{\mu\to i}^t\right)$$

$V]$,

$] - f_w f_w^\intercal$.

$k\times k$

$$B_{\mu\to i}^t\equiv\frac{X_{\mu i}}{\sqrt{n}}g_{\text{out}}(\omega_{i\mu}^t,Y_\mu,V_{i\mu}^t),$$

$$A_{\mu\to i}^t\equiv-\frac{X_{\mu i}^2}{n}\partial_\omega g_{\text{out}}(\omega_{i\mu}^t,Y_\mu,V_{i\mu}^t)$$

$$,Y_\mu,V_{i\mu}^t)+\frac{1}{2}\frac{X_{\mu i}^2}{n}W_i^\intercal g_{\text{out}}g_{\text{out}}^\intercal(\omega_{i\mu}^t,Y_\mu,V_{i\mu}^t)W_i+$$

why omit?

$$i\}=\frac{1}{\mathcal{Z}_{\mu\to i}}\left\{1+W_i^\intercal B_{\mu\to i}^t+\frac{1}{2}W_i^\intercal B_{\mu\to i}^t(B_{\mu\to i}^t)^\intercal(W_i)-\frac{1}{2}W_i^\intercal A_{\mu\to i}^tW_i\right\}$$

$$=\sqrt{\frac{\det(A_{\mu\to i}^t)}{(2\pi)^K}}\exp\left(-\frac{1}{2}\left(W_i^\intercal-(A_{\mu\to i}^t)^{-1}B_{\mu\to i}^t\right)^\intercal A_{\mu\to i}^t\left(W_i^\intercal-(A_{\mu\to i}^t)^{-1}B_{\mu\to i}^t\right)\right)$$

$$m_{\mu\to i}\left(t,x_i\right)=\sqrt{\frac{A_{\mu\to i}^t}{2\pi N}}\exp\left\{-\frac{x_i^2}{2N}A_{\mu\to i}^t+B_{\mu\to i}^t\frac{x_i}{\sqrt{N}}-\frac{\left(B_{\mu\to i}^t\right)^2}{2A_{\mu\to i}^t}\right\}$$

AMP

$$m_{i\to\mu}(w_i) \propto P_0(w_i)\, e^{-\frac{\left(w_i - R_{i\to\mu}\right)^2}{2\Sigma_{i\to\mu}}}$$

define

$$f_{\mathrm{w}} \equiv \frac{\int \mathrm{d}w\, w P_0(w) e^{-\frac{(w-R)^2}{2\Sigma}}}{\int \mathrm{d}w\, P_0(w) e^{-\frac{(w-R)^2}{2\Sigma}}}$$

$$\Sigma^t_{i\to\mu} = \left(\sum_{\nu\neq\mu} A^t_{\nu\to i}\right)^{-1}$$

$$R^t_{i\to\mu} = \Sigma^t_{i\to\mu}\left(\sum_{\nu\neq\mu} B^t_{\nu\to i}\right)$$

$$\hat{w}_{i\to\mu} \equiv \int \mathrm{d}w_i\, m_{i\to\mu}(w_i)\, w_i$$

$$v_{i\to\mu} \equiv \int \mathrm{d}w_i\, m_{i\to\mu}(x_i)\, w_i^2 - \hat{w}^2_{i\to\mu}$$

⇒

$$\hat{w}_{\mu\to i} = f_{\mathrm{w}}(\Sigma, R)$$

$$v_{\mu\to i} = \partial_R f_{\mathrm{w}}(\Sigma, R)$$

r-BP Algorithm

Algorithm 1 relaxed Belief-Propagation (r-BP)

Input: **y**
Initialize: $\mathbf{a}_{i\to\mu}(t=0), \mathbf{v}_{i\to\mu}(t=0)$, t = 1
repeat
r-BP Update of $\{\omega_{\mu\to i}, V_{\mu\to i}\}$

$$V_{\mu\to i}(t) \leftarrow \sum_{j\neq i} F^2_{\mu j} v_{j\to\mu}(t-1)$$

$$\omega_{\mu\to i}(t) \leftarrow \sum_{j\neq i} F_{\mu j} a_{j\to\mu}(t-1)$$

r-BP Update of $\{A_{\mu\to i}, B_{\mu\to i}\}$

$$B_{\mu\to i}(t) \leftarrow g_{\mathrm{out}}(\omega_{\mu\to i}(t), y_\mu, V_{\mu\to i}(t))\, F_{\mu i}\,,$$
$$A_{\mu\to i}(t) \leftarrow -\partial_\omega g_{\mathrm{out}}(\omega_{\mu\to i}(t), y_\mu, V_{\mu\to i}(t)) F^2_{\mu i}$$

r-BP Update of $\{R_{\mu\to i}, \Sigma_{\mu\to i}\}$

$$\Sigma_{i\to\mu}(t) \leftarrow \frac{1}{\sum_{\nu\neq\mu} A_{\nu\to i}(t)}$$

$$R_{i\to\mu}(t) \leftarrow \Sigma_{i\to\mu}(t) \sum_{\nu\neq\mu} B_{\nu\to i}(t)$$

AMP Update of the estimated partial marginals $\mathbf{a}_{i\to\mu}(t), \mathbf{v}_{i\to\mu}(t)$

$$a_{i\to\mu}(t) \leftarrow f_a\left(\Sigma_{i\to\mu}(t), R_{i\to\mu}(t)\right)\,,$$
$$v_{i\to\mu}(t) \leftarrow f_v\left(\Sigma_{i\to\mu}(t), R_{i\to\mu}(t)\right)\,.$$

t ← t + 1
until Convergence on $\mathbf{a}_{i\to\mu}(t), \mathbf{v}_{i\to\mu}(t)$
output: Estimated marginals mean and variances:

$$a_i \leftarrow f_a\left(\frac{1}{\sum_\nu A_{\nu\to i}}, \frac{\sum_\nu B_{\nu\to i}}{\sum_\nu A_{\nu\to i}}\right)\,,$$
$$v_i \leftarrow f_v\left(\frac{1}{\sum_\nu A_{\nu\to i}}, \frac{\sum_\nu B_{\nu\to i}}{\sum_\nu A_{\nu\to i}}\right)\,.$$

AMP

$\boxed{XXX}_{\mu\to i} \longrightarrow \boxed{XXX}_{\mu}$

r-BP ⟶ AMP

$$\omega_{\mu\to i} = \sum_{j\neq i} X_{\mu j}\hat{w}_{j\to\mu} \longrightarrow \omega_{\mu} = \sum_{i} X_{\mu i}\hat{w}_{i\to\mu}$$

$$V_{\mu\to i} = \sum_{j\neq i} X_{\mu j}^2 v_{j\to\mu} \qquad V_{\mu} = \sum_{i} X_{\mu i}^2 v_{i\to\mu}$$

$\boxed{XXX}_{\mu\to i} \to \boxed{XXX}_{\mu}$ ⟶

$$\Sigma_{\mu}^{t+1} = \frac{1}{\sum_{\mu} A_{\mu\to i}^{t+1}}$$

$$R_{\mu}^{t+1} = \frac{\sum_{\mu} B_{\mu\to i}^{t+1}}{\sum_{\mu} A_{\mu\to i}^{t}}$$

$$V_{\mu}^{t+1} = \sum_{i} X_{\mu i}^2 v_{i\to\mu}^t \approx \sum_{i} X_{\mu i}^2 v_i^t$$

$$\omega_{\mu\to i} = \sum_{j\neq i} X_{\mu j}\hat{w}_{j\to\mu}$$

$$B_{\mu\to i}^t = X_{\mu i}\, f_{\text{out}}\left(\omega_{\mu\to i}^t, y_\mu, V_{\mu\to i}^t\right)$$

$$A_{\mu\to i}^t = -X_{\mu i}^2\, \partial_\omega f_{\text{out}}\left(\omega_{\mu\to i}^t, y_\mu, V_{\mu\to i}^t\right)$$

⟶

$$(\Sigma_i)^{t+1} \approx \left[-\sum_{\mu} X_{\mu i}^2 \partial_\omega f_{\text{out}}\left(\omega_{\mu}^{t+1}, y_\mu, V_{\mu}^{t+1}\right)\right]^{-1}$$

$$R_i^{t+1} = \left[-\sum_{\mu} X_{\mu i}^2 \partial_\omega f_{\text{out}}\left(\omega_{\mu}^{t+1}, y_\mu, V_{\mu}^{t+1}\right)\right]^{-1} \times \left[\sum_{\mu} X_{\mu i} f_{\text{out}}\left(\omega_{\mu\to i}^{t+1}, y_\mu, V_{\mu\to i}^{t+1}\right)\right]$$

AMP

- $$f_{\text{out}}\left(\omega_{\mu\to i}^{t+1}, y_\mu, V_{\mu\to i}^{t+1}\right)$$

$$\approx f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right) - X_{\mu i}\hat{w}_{i\to\mu}^t \partial_\omega f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)$$

$$\approx f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right) - X_{\mu i}\hat{w}_i^t \partial_\omega f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)$$

- $$R_i^{t+1}$$

$$= (\Sigma_i)^{t+1} \times \left[\sum_\mu X_{\mu i} f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right) - X_{\mu i}^2 \hat{w}_i^t \partial_\omega f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)\right]$$

$$= \hat{w}_i^t + (\Sigma_i)^{t+1} \sum_\mu X_{\mu i} f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)$$

- $$\hat{w}_{i\to\mu}^t = f_{\text{w}}\left(R_{i\to\mu}^t, \Sigma_{i\to\mu}\right) \approx f_{\text{w}}\left(R_{i\to\mu}^t, \Sigma_i\right)$$

$$\approx f_{\text{w}}\left(R_i^t, \Sigma_i\right) - B_{\mu\to i}^t \partial_R f_{\text{w}}\left(R_i^t, \Sigma_i\right)$$

$$\approx \hat{w}_i^t - f_{\text{out}}\left(\omega_\mu^t, y_\mu, V_\mu^t\right) X_{\mu i} v_i^t$$

$$\hat{w}_i^{t+1} = \frac{\Sigma_i^{t+1}}{1 + R_i^{t+1}}$$

- $$\omega_\mu^{t+1} = \sum_i X_{\mu i}\hat{w}_i^t - \sum_i f_{\text{out}}\left(\omega_\mu^t, y_\mu, V_\mu^t\right) X_{\mu i}^2 v_i^t$$

$$= \sum X_{\mu i}\hat{w}_i - V_\mu^t f_{\text{out}}\left(\omega_\mu^t, y_\mu, V_\mu^t\right)$$

$$v_i^{t+1} = \frac{1}{1 + R_i^{t+1}}$$

输入：y 、X

初始化 $\hat{w}^0$ 、v^0 、f_{out}^0

设置迭代次数 $t = 1$，最大迭代次数 T

while $t < T$ **do**

计算 ω 、V

$$\omega_\mu^{t+1} \leftarrow \frac{1}{\sqrt{n}} \sum_i X_{\mu i}\hat{w}_i^t - V_\mu^t f_{\text{out}}^t\left(\omega_\mu^t, y_\mu, V_\mu^t\right)$$

$$V_\mu^{t+1} \leftarrow \frac{1}{n} \sum_i X_{\mu i}^2 v_i^t$$

计算 f_{out}

$$f_{\text{out}}^{t+1} \leftarrow f_{\text{out}}(y, \omega^{t+1}, V^{t+1})$$

计算 Σ 、R

$$\Sigma_i^{t+1} \leftarrow \left[-\frac{1}{n} \sum_\mu X_{\mu i}^2 \partial_\omega f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)\right]^{-1}$$

$$R_i^{t+1} \leftarrow \hat{w}_i^t + \frac{1}{\sqrt{n}} (\Sigma_i)^{t+1} \sum_\mu X_{\mu i} f_{\text{out}}\left(\omega_\mu^{t+1}, y_\mu, V_\mu^{t+1}\right)$$

计算 $\hat{w}$ 、v

$$\hat{w}_i^{t+1} \leftarrow \frac{\Sigma_i^{t+1}}{1 + R_i^{t+1}}$$

$$v_i^{t+1} \leftarrow \frac{1}{1 + R_i^{t+1}}$$

if $\hat{w}$ 、v 不再变化 **then**

停止迭代

end

$t \leftarrow t + 1$

SE

模型出发点：

$$\omega_{\mu\to i} = \frac{1}{\sqrt{n}} \sum_{j\neq i} X_{\mu i}\hat{w}_i$$

$$z_{\mu\to i} = \frac{1}{\sqrt{n}} \sum_{j\neq i} X_{\mu i} w^*_{i\to j}$$

$$V_\mu = \frac{1}{n} \sum_i X_{\mu i}^2 v_i$$

算法出发点：

$$\Sigma_i = \frac{1}{\sum_\mu A_{\mu\to i}}$$

$$R_i = \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}$$

$$f_{\rm w}(\Sigma, R) \equiv \frac{\int {\rm d}w\, w P_0(w) e^{-\frac{(w-R)^2}{2\Sigma}}}{\int {\rm d}w P_0(w) e^{-\frac{(w-R)^2}{2\Sigma}}}$$

$$f_{\rm out}(\omega, y, V) \equiv \frac{\int {\rm d}z P_{\rm out}(y|z)(z-\omega) e^{-\frac{(z-\omega)^2}{2V}}}{V \int {\rm d}z P_{\rm out}(y|z) e^{-\frac{(z-\omega)^2}{2V}}}$$

$$\begin{aligned}\frac{R_i}{\Sigma_i} &= \sum_\mu B_{\mu\to i} \\ &= \sum_\mu X_{\mu i} f_{\rm out}(\omega_{\mu\to i}, y_\mu, V_{\mu\to i}) \\ &= \sum_\mu X_{\mu i} f_{\rm out}(\omega_{\mu\to i}, {\rm sign}[\sum_{j\neq i} X_{\mu j} w_j^* + X_{\mu i} w_i^*], V) \\ &= \sum_\mu X_{\mu i} f_{\rm out}(\omega_{\mu\to i}, {\rm sign}[\sum_{j\neq i} X_{\mu j} w_j^*], V) + \sum_\mu X_{\mu i} f_{\rm out}(\omega_{\mu\to i}, {\rm sign}[X_{\mu i} w_i^*], V)\end{aligned}$$

定义以下序参量：

$$\hat{q} = \alpha \mathbb{E}_{\omega,z}[f_{\rm out}^2(\omega, {\rm sign}[z], V)]$$

$$\hat{m} = \alpha \mathbb{E}_{\omega,z}[\partial_z f_{\rm out}(\omega, {\rm sign}[z], V]$$

⟶

$$\frac{R_i}{\Sigma_i} = \mathcal{N}(0,1)\sqrt{\hat{q}} + w_i^* \hat{m}$$

闭合方程：

$$q = \mathbb{E}_{w^*}\mathbb{E}_{R,\Sigma}[f_{\rm w}^2(\Sigma, R)]$$

$$m = \mathbb{E}_{w^*}\mathbb{E}_{R,\Sigma}[w^* f_{\rm w}(\Sigma, R)]$$

⟵

$$q = \mathbb{E}[\omega^2] = \mathbb{E}[\hat{w}^2]$$

$$m = \mathbb{E}[z\omega] = \mathbb{E}[w^* \hat{w}]$$

SE

$$q = \mathbb{E}_{w^*}\mathbb{E}_{R,\Sigma}[f_{\mathrm{w}}^2(\Sigma, R)]$$

$$m = \mathbb{E}_{w^*}\mathbb{E}_{R,\Sigma}[w^* f_{\mathrm{w}}(\Sigma, R)]$$

$$\hat{q} = \alpha\mathbb{E}_{\omega,z}[f_{\mathrm{out}}^2(\omega, \mathrm{sign}[z], V)]$$

$$\hat{m} = \alpha\mathbb{E}_{\omega,z}[\partial_z f_{\mathrm{out}}(\omega, \mathrm{sign}[z], V]$$

Nishimori condition:q=m

Bayes optimal

$$q = m = \mathbb{E}_{w^*}\mathbb{E}_{R,\Sigma}[w^* f_{\mathrm{w}}(\Sigma, R)]$$

$$\hat{q} = \hat{m} = \alpha\mathbb{E}_{\omega,z}[\partial_z f_{\mathrm{out}}(\omega, \mathrm{sign}[z], V]$$

显式表达：

$$q^{t+1} = \int dx P_X(x) \int d\xi \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} f_{w^*}^2(\frac{1}{\hat{q}^t}, x + \frac{\xi}{\sqrt{\hat{q}^t}})$$

$$\hat{q}^t = -\int dp \int dz \frac{e^{-\frac{p^2}{2m^t}} e^{-\frac{(z-p)^2}{2(1-m^t)}}}{2\pi\sqrt{m^t(1-m^t)}} \partial_p f_{\mathrm{out}}(p, \mathrm{sign}[z], 1-m^t)$$

简单并不断重复：

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} x e^{-x^2} dx = 0$$

$$\int_{-\infty}^{\infty} x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

$$q = \frac{\hat{q}}{1+\hat{q}}$$

$$\hat{q} = \frac{2}{\pi}\frac{\alpha}{1-q}\int D\xi \frac{\exp\left\{-\frac{q\xi^2}{1-q}\right\}}{1+\mathrm{erf}\left(\frac{\sqrt{q}\xi}{\sqrt{2(1-q)}}\right)}$$

same as replica

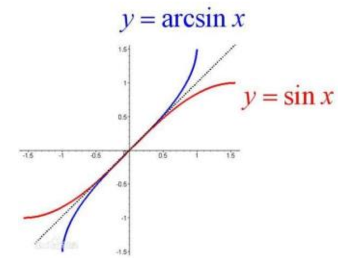
Asymptotic solution $\alpha \to \infty,\ q \to 1$

$$\int D\xi \frac{e^{-\frac{q_b \xi^2}{1-q_b}}}{\left(1+\mathrm{erf}\left(\frac{\sqrt{q_b}\xi}{\sqrt{2(1-q_b)}}\right)\right)} = \int \mathrm{d}\xi \frac{\frac{-e^{\frac{\xi^2(q_b+1)}{2(1-q_b)}}}{\sqrt{2\pi}}}{\left(1+\mathrm{erf}\left(\frac{\sqrt{q_b}\xi}{\sqrt{2(1-q_b)}}\right)\right)} \simeq \int \mathrm{d}\xi \frac{\frac{-e^{\frac{\xi^2}{1-q_b}}}{\sqrt{2\pi}}}{\left(1+\mathrm{erf}\left(\frac{\xi}{\sqrt{2(1-q_b)}}\right)\right)}$$

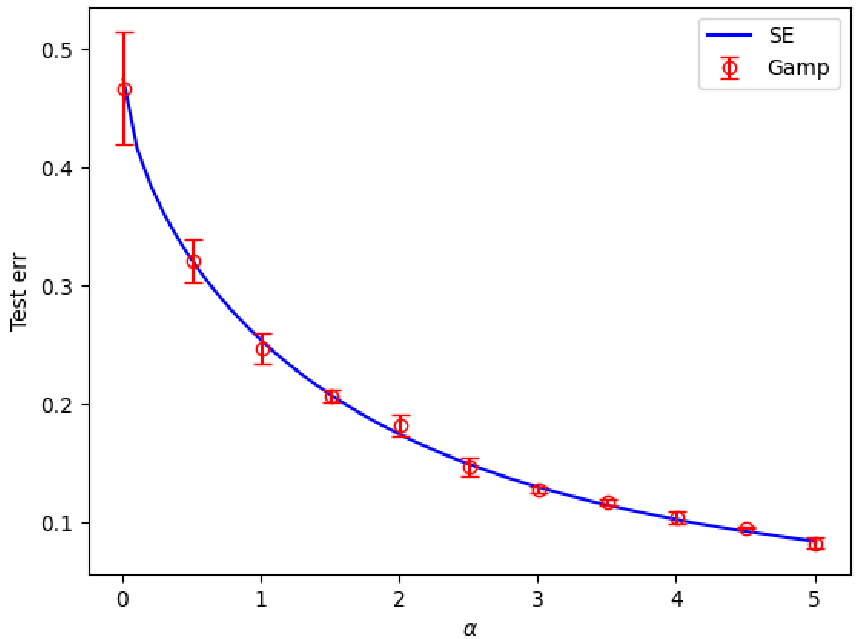
$$= \frac{\sqrt{1-q_b}}{\sqrt{2\pi}} \int \mathrm{d}\eta \frac{e^{-\eta^2}}{1+\mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)} = \frac{c_0}{\sqrt{2\pi}}\sqrt{1-q_b}$$

$$c_0 \equiv \int \mathrm{d}\eta \frac{e^{-\eta^2}}{1+\mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)} \simeq 2.83748$$

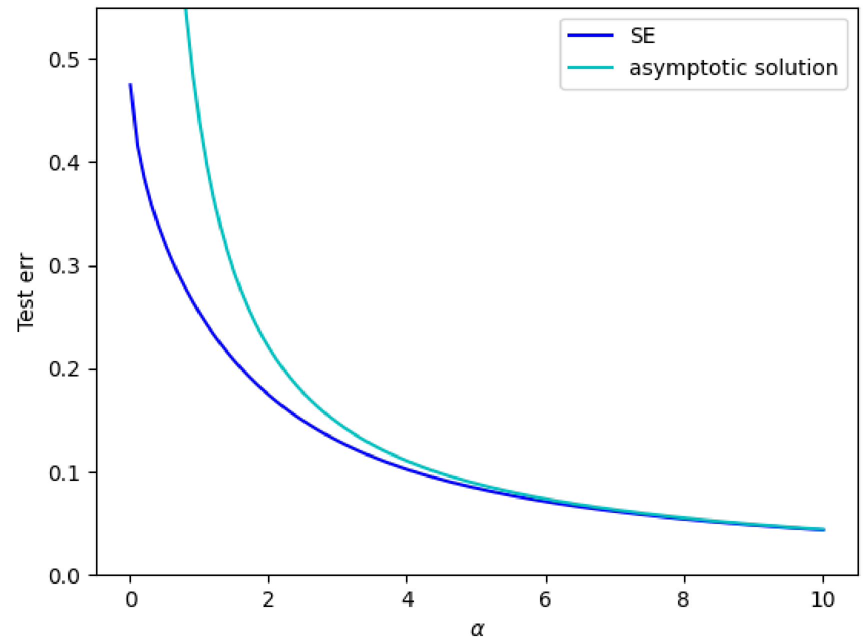
$$q_b = \frac{1}{2}\left(\alpha k\sqrt{\alpha^2 k^2 + 4} - \alpha^2 k^2\right) \underset{\alpha\to\infty}{\simeq} 1 - \frac{1}{\alpha^2 k^2} \qquad k \equiv \frac{2c_0}{\pi\sqrt{2\pi}} \simeq 0.720647$$



Conclusion



$$\varepsilon_g^{bayes}(\alpha) = \frac{1}{\pi}\arccos\left(\sqrt{q^T}\right)$$



$$\varepsilon_g^{bayes}(\alpha) \underset{\alpha\to\infty}{\simeq} \frac{1}{k\pi}\frac{1}{\alpha} \simeq \frac{0.4417}{\alpha}$$

Replica

$$P(\mathbf{w}|\mathbf{y},\mathbf{X}) = \frac{P(\mathbf{y}|\mathbf{X},\mathbf{w})P(\mathbf{w})}{P(\mathbf{y},\mathbf{X})} = \frac{1}{\mathcal{Z}}P(\mathbf{y}|\mathbf{X},\mathbf{w})P(\mathbf{w})$$

partition function

$$\mathcal{Z}(\mathbf{y},\mathbf{X}) = P(\mathbf{y},\mathbf{X}) = \int \mathrm{d}\mathbf{w}\, P(\mathbf{w},\mathbf{y},\mathbf{X}) = \int \mathrm{d}\mathbf{w}\, P(\mathbf{y}|\mathbf{w},\mathbf{X})P(\mathbf{w})$$

$$= \int \mathrm{d}\mathbf{z}\, P(\mathbf{y}|\mathbf{z}) \int \mathrm{d}\mathbf{w}\, P(\mathbf{w})\,\delta\left(\mathbf{z} - \frac{1}{\sqrt{n}}\mathbf{w}\mathbf{X}\right)$$

$$\Phi = \frac{1}{n}\mathbb{E}_{\mathbf{y},\mathbf{X}}\ \log \mathcal{Z}(\mathbf{y},\boldsymbol{X}) \xrightarrow{\text{Replica}} \Phi = \frac{1}{n}\lim_{r\to 0}\frac{\partial \log \mathbb{E}_{\mathbf{y},\mathbf{X}}\ \left[\mathcal{Z}(\mathbf{y},\boldsymbol{X})^r\right]}{\partial r}$$

$$\mathbb{E}_{\mathbf{y},\mathrm{X}}\left[\mathcal{Z}(\mathbf{y},\mathrm{X})^r\right] = \mathbb{E}_{\mathbf{w}^*,\mathrm{X}}\left[\prod_{a=1}^{r}\int_{\mathbb{R}^n} \mathrm{d}\mathbf{z}^a P_{\mathrm{out}^a}\left(\mathbf{y} \mid \mathbf{z}^a\right)\int_{\mathbb{R}^d}\mathrm{d}\mathbf{w}^a P_{\mathbf{w}^a}\left(\mathbf{w}^a\right)\delta\left(\mathbf{z}^a - \frac{1}{\sqrt{d}}\mathrm{X}\mathbf{w}^a\right)\right]$$

$$= \mathbb{E}_{\mathrm{X}}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{y}P(\mathbf{y}\mid X)\int_{\mathbb{R}^n}\mathrm{d}\mathbf{z}^a P_{\mathrm{out}^a}\left(\mathbf{y}\mid\mathbf{z}^a\right)\int_{\mathbb{R}^d}\mathrm{d}\mathbf{w}^a P_{\mathbf{w}^a}\left(\mathbf{w}^a\right)\delta\left(\mathbf{z}^a - \frac{1}{\sqrt{d}}\mathrm{X}\mathbf{w}^a\right)$$

$$= \mathbb{E}_{\mathrm{X}}\boxed{\int_{\mathbb{R}^n}\mathrm{d}\mathbf{y}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{z}^* P_{\mathrm{out}*}\left(\mathbf{y}\mid\mathbf{z}^*\right)\int_{\mathbb{R}^d}\mathrm{d}\mathbf{w}^* P_{\mathbf{w}*}(\mathbf{w}^*)\,\delta\left(\mathbf{z}^* - \frac{1}{\sqrt{d}}\mathbf{X}^*\right)}$$

Average over y $P(\mathbf{y} \mid X)$

$$\times\left[\prod_{a=1}^{r}\int_{\mathbb{R}^n}\mathrm{d}^a P_{\mathrm{out}^a}\left(\mathbf{y}\mid\mathbf{z}^a\right)\int_{\mathbb{R}^d}\mathrm{d}\mathbf{w}^a P_{\mathbf{w}^a}\left(\mathbf{w}^a\right)\delta\left(\mathbf{z}^a - \frac{1}{\sqrt{d}}\mathrm{X}\mathbf{w}^a\right)\right]$$

$$= \mathbb{E}_{\mathrm{X}}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{y}\prod_{a=0}^{r}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{z}^a P_{\mathrm{out}^a}\left(\mathbf{y}\mid\mathbf{z}^a\right)\int_{\mathbb{R}^d}\mathrm{d}\mathbf{w}^a P_{\mathbf{w}^a}\left(\mathbf{w}^a\right)\delta\left(\mathbf{z}^a - \frac{1}{\sqrt{d}}\mathrm{X}\mathbf{w}^a\right)$$

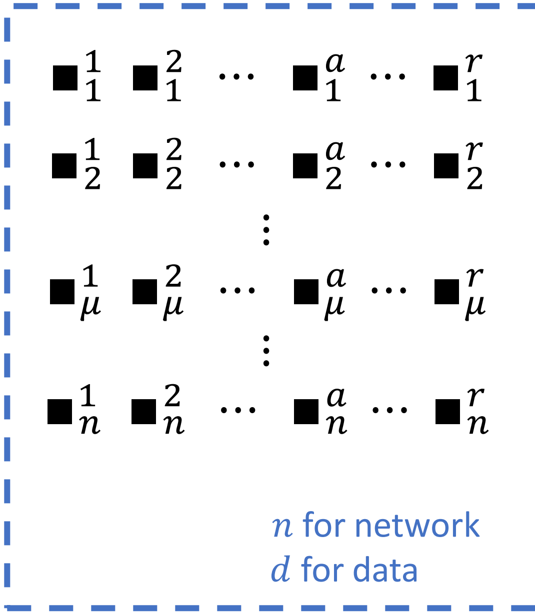
Replica

Assuming X is i.i.d, according to the central limit theorem

$$z_\mu^a = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i^{(\mu)} w_i^a \sim \mathcal{N}\left(\mathbb{E}_{\mathrm{X}}\left[z_\mu^a\right], \mathbb{E}_{\mathrm{X}}\left[z_\mu^a z_\mu^b\right]\right) \qquad a,b=0..r$$

$$\mathbb{E}_{\mathrm{X}}\left[z_\mu^a\right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}_{\mathrm{X}}\left[x_i^{(\mu)}\right] w_i^a = 0$$

$$\mathbb{E}_{\mathrm{X}}\left[z_\mu^a z_\mu^b\right] = \frac{1}{n} \sum_{ij} \mathbb{E}_{\mathrm{X}}\left[x_i^{(\mu)} x_j^{(\mu)}\right] w_i^a w_j^b = \frac{1}{n} \sum_{ij} \delta_{ij} w_i^a w_j^b = \frac{1}{n} \mathbf{w}^a \cdot \mathbf{w}^b \equiv Q$$



The integration of column vectors → row vectors.

$$\tilde{\mathbf{z}}_\mu \equiv \left(z_\mu^a\right)_{a=0\cdots r}, \quad \tilde{\mathbf{w}}_i \equiv \left(w_i^a\right)_{a=0\cdots r}$$

$$\tilde{\mathbf{z}}_\mu \sim P_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}}; Q) = \mathcal{N}_{\tilde{\mathbf{z}}}\left(\mathbf{0}_{r+1}, Q\right) \qquad P_{\tilde{\mathrm{w}}}(\tilde{\mathbf{w}}) = \prod_{a=0}^{r} P_{\mathrm{w}}\left(\tilde{w}^a\right)$$

$$\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathcal{Z}(\mathbf{y},\mathbf{X})^r\right] = \mathbb{E}_{\mathbf{X}} \int \mathrm{d}\mathbf{y} \prod_{a=0}^{r} \int \mathrm{d}\mathbf{z}^a P_{\mathrm{out}^a}\left(\mathbf{y} \mid \mathbf{z}^a\right) \int \mathrm{d}\mathbf{w}^a P_{\mathbf{w}^a}\left(\mathbf{w}^a\right) \delta\left(\mathbf{z}^a - \frac{1}{\sqrt{d}} \mathbf{X}\mathbf{w}^a\right)$$

$$= \left[\int \mathrm{d}y \int \mathrm{d}\tilde{\mathbf{z}} P_{\mathrm{out}}(y \mid \tilde{\mathbf{z}}) P_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}}; Q(\tilde{\mathbf{w}}))\right]^d \left[\int \mathrm{d}\tilde{\mathbf{w}} P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}})\right]^n$$

Replica

Introduce Fourier transform of Dirac δ function

$$1 = \int \mathrm{d}Q \prod_{a\le b} \delta\left(nQ_{ab} - \sum_{i=1}^{n} w_i^a w_i^b\right)$$

$$\propto \int \mathrm{d}Q \int \mathrm{d}\hat{Q} \prod_{a\le b} \exp\left\{-\hat{Q}_{ab}\left(nQ_{ab} - \sum_{i=1}^{n} w_i^a w_i^b\right)\right\}$$

$$\propto \int \mathrm{d}Q \int \mathrm{d}\hat{Q} \exp\left\{-\sum_{a\le b}\hat{Q}_{ab}\left(nQ_{ab} - \sum_{i=1}^{n} w_i^a w_i^b\right)\right\}$$

$$\propto \int \mathrm{d}Q \int \mathrm{d}\hat{Q} \exp(-n\,\mathrm{Tr}[Q\hat{Q}]) \exp\left(\frac{1}{2}\sum_{i=1}^{n} \tilde{\mathbf{w}}_i^\top \hat{Q} \tilde{\mathbf{w}}_i\right)$$

$$\delta(x) = \frac{1}{2i\pi}\int_{i\mathbb{R}} \mathrm{d}\hat{x} e^{-\hat{x}x}$$

$$tr(\mathbf{AB}) = \textstyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}$$

$a > b:$

$$\delta^2(x) = \frac{b-a}{2\pi}\delta(x), \qquad x \in [a,b]$$

Insert it into the partition function

$$\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathcal{Z}(\mathbf{y},\mathbf{X})^r\right] \propto \int \mathrm{d}Q \int \mathrm{d}\hat{Q}\ \exp(-n\,\mathrm{Tr}[Q\hat{Q}]) \exp\left(\frac{1}{2}\sum_{i=1}^{n} \tilde{\mathbf{w}}_i^\top \hat{Q} \tilde{\mathbf{w}}_i\right)$$
$$\left[\int \mathrm{d}y \int \mathrm{d}\tilde{\mathbf{z}}\ P_{\mathrm{out}}(y \mid \tilde{\mathbf{z}}) P_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}}; Q(\tilde{\mathbf{w}}))\right]^d \left[\int \mathrm{d}\tilde{\mathbf{w}} P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}})\right]^n$$
$$\propto \iint \mathrm{d}Q\,\mathrm{d}\hat{Q}\ e^{n\Phi^{(r)}(Q,\hat{Q})}$$

where

$$\Phi^{(r)}(Q,\hat{Q}) = -\mathrm{Tr}[Q\hat{Q}] + \log \Psi_{\mathrm{w}}^{(r)}(\hat{Q}) + \alpha \log \Psi_{\mathrm{out}}^{(r)}(Q)$$

$$\Psi_{\mathrm{w}}^{(r)}(\hat{Q}) = \int \mathrm{d}\tilde{\mathbf{w}}\ P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}\hat{Q}\tilde{\mathbf{w}}}$$

$$\Psi_{\mathrm{out}}^{(r)}(Q) = \int \mathrm{d}y \int \mathrm{d}\tilde{\mathbf{z}}\ P_{\tilde{z}}(\tilde{\mathbf{z}}; Q) P_{\mathrm{out}}(y \mid \tilde{\mathbf{z}})$$

$$\Phi = \frac{1}{n}\lim_{r\to 0}\frac{\partial \log \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathcal{Z}(\mathbf{y},\boldsymbol{X})^r\right]}{\partial r}$$

Laplace approximation

$$\Phi(\alpha) = \mathrm{extr}_{Q,\hat{Q}}\left\{\lim_{r\to 0}\frac{\partial \Phi^{(r)}(Q,\hat{Q})}{\partial r}\right\}$$

Replica

Replica symmetric ansatz

$$Q_{\text{rs}} = \begin{pmatrix} Q^0 & m & \dots & m \\ m & Q & \dots & \dots \\ \dots & \dots & \dots & q \\ m & \dots & q & Q \end{pmatrix} \quad \text{and} \quad \hat{Q}_{\text{rs}} = \begin{pmatrix} \hat{Q}^0 & \hat{m} & \dots & \hat{m} \\ \hat{m} & -\frac{1}{2}\hat{Q} & \dots & \dots \\ \dots & \dots & \dots & \hat{q} \\ \hat{m} & \dots & \hat{q} & -\frac{1}{2}\hat{Q} \end{pmatrix}$$

$$m = \frac{1}{n}\mathbf{w}^a \cdot \mathbf{w}^\star$$

$$q = \frac{1}{n}\mathbf{w}^a \cdot \mathbf{w}^b$$

$$Q = \frac{1}{n}||\mathbf{w}^a||_2^2$$

$$Q^0 = \rho_{\mathbf{w}^\star} = \frac{1}{n}||\mathbf{w}^\star||_2^2$$

$$\Phi(\alpha) = \text{extr}_{Q,\hat{Q}} \left\{ \lim_{r\to 0} \frac{\partial \Phi^{(r)}(Q,\hat{Q})}{\partial r} \right\}$$

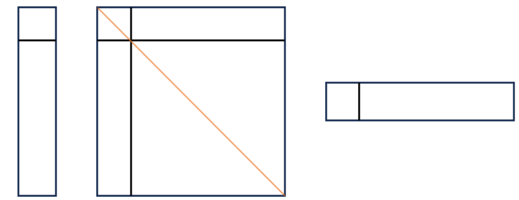
$$\Phi^{(r)}(Q,\hat{Q}) = -\operatorname{Tr}[Q\hat{Q}] + \log \Psi_{\text{w}}^{(r)}(\hat{Q}) + \alpha \log \Psi_{\text{out}}^{(r)}(Q)$$

- $\operatorname{Tr}\left(Q\hat{Q}\right)\Big|_{\text{rs}} = Q^0\hat{Q}^0 + rm\hat{m} - \frac{1}{2}rQ\hat{Q} + \frac{r(r-1)}{2}q\hat{q}$

$$tr(\mathbf{AB}) = \textstyle\sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ji}$$

$$\begin{aligned}
&\text{extr}\left\{ \lim_{r\to 0} \frac{\partial}{\partial r} \left(-\operatorname{Tr}[Q\hat{Q}] \right) \right\} \\
&= \text{extr}\left\{ \lim_{r\to 0} \frac{\partial}{\partial r} \left(-Q^0\hat{Q}^0 - rm\hat{m} + \frac{1}{2}rQ\hat{Q} - \frac{r(r-1)}{2}q\hat{q} \right) \right\} \\
&= \text{extr}\left\{ \lim_{r\to 0} \left(-m\hat{m} + \frac{1}{2}Q\hat{Q} - \left(r - \frac{1}{2}\right) q\hat{q} \right) \right\} \\
&= \text{extr}\left\{ -m\hat{m} + \frac{1}{2}Q\hat{Q} + \frac{1}{2}q\hat{q} \right\}
\end{aligned}$$

Replica



- $$\Psi_{\mathrm{w}}^{(r)}(\hat{Q}) = \int \mathrm{d}\tilde{\mathbf{w}}\ P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}\hat{Q}\tilde{\mathbf{w}}}$$

$$\tilde{\mathbf{w}}\hat{Q}_{\mathrm{rs}}\tilde{\mathbf{w}} = w^\star \hat{Q}^0 w^\star + 2\sum_{a=1}^{r} w^\star \hat{m} w^a - \left(\hat{Q}+\hat{q}\right)\sum_{a=1}^{r}(w^a)^2 + \hat{q}\left(\sum_{a=1}^{r} w^a\right)^2$$

$$\Psi_{\mathrm{w}}^{(r)}(\hat{Q})\Big|_{\mathrm{rs}} = \int \mathbf{d}\tilde{\mathbf{w}} P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}\hat{Q}_{\mathrm{rs}}\tilde{\mathbf{w}}}$$
$$= \mathbb{E}_{w^\star} e^{\frac{1}{2}\hat{Q}^0(w^\star)^2} \int \mathbf{d}\tilde{\mathbf{w}} P_{\tilde{\mathbf{w}}}(\tilde{\mathbf{w}}) e^{w^\star \hat{m}\sum_{a=1}^{r} w^a - \frac{1}{2}(\hat{Q}+\hat{q})\sum_{a=1}^{r}(\tilde{w}^a)^2 + \frac{1}{2}\hat{q}\left(\sum_{a=1}^{r} w^a\right)^2}$$
$$= \mathbb{E}_{\xi,w^\star} e^{\frac{1}{2}\hat{Q}^0(w^\star)^2}\left[\mathbb{E}_w \exp\left(\hat{m}w^\star w - \frac{1}{2}(\hat{Q}+\hat{q})w^2 + \hat{q}^{1/2}\xi w\right)\right]^r$$

H-S transformation

$$\mathbb{E}_\xi \exp(\sqrt{a}\xi) = e^{\frac{a}{2}}$$

$$\Psi_{\mathrm{w}}(\hat{Q}) = \mathrm{extr}\left\{\lim_{r\to 0}\frac{\partial}{\partial r}\left(\log \Psi_{\mathrm{w}}^{(r)}(\hat{Q})\right)\right\}$$
$$= \mathrm{extr}\left\{\lim_{r\to 0}\frac{\partial}{\partial r}\mathbb{E}_{\xi,w^\star} r \log\left[\mathbb{E}_w \exp\left(\hat{m}w^\star w - \frac{1}{2}(\hat{Q}+\hat{q})w^2 + \hat{q}^{1/2}\xi w\right)\right]\right\}$$
$$= \mathrm{extr}\left\{\mathbb{E}_{\xi,w^\star}\log\left[\mathbb{E}_w \exp\left(\hat{m}w^\star w - \frac{1}{2}(\hat{Q}+\hat{q})w^2 + \hat{q}^{1/2}\xi w\right)\right]\right\}$$
$$= \mathrm{extr}\left\{\mathbb{E}_{\xi,w^\star}\exp\left(-\frac{1}{2}\hat{q}^{-1}\hat{m}^2(w^\star)^2 + \xi\hat{q}^{-\frac{1}{2}}\hat{m}w^\star\right)\log\left[\mathbb{E}_w \exp\left(-\frac{1}{2}(\hat{Q}+\hat{q})w^2 + \hat{q}^{1/2}\xi w\right)\right]\right\}$$

decouple the teacher and student expectations

$$\xi \leftarrow \xi + \hat{q}^{-\frac{1}{2}}\hat{m}w^\star$$

Replica

$$\mathbb{E}_{\xi,w^\star} \exp\left(-\frac{1}{2}\hat{q}^{-1}\hat{m}^2(w^\star)^2 + \xi\hat{q}^{-\frac{1}{2}}\hat{m}w^\star\right) \log\left[\mathbb{E}_w \exp\left(-\frac{1}{2}(\hat{Q}+\hat{q})w^2 + \hat{q}^{1/2}\xi w\right)\right]$$

define $\mathcal{Z}_{\mathrm{w}}(\gamma, \Lambda) \equiv \mathbb{E}_{w\sim P_{\mathrm{w}}}\left[e^{-\frac{1}{2}\Lambda w^2+\gamma w}\right] \xrightarrow{P_{\mathrm{w}} \sim \mathcal{N}(0,1)} \frac{e^{\frac{\gamma^2}{2(\Lambda+1)}}}{\sqrt{\Lambda+1}}$

Bayesian optimization $\mathcal{Z}_{\mathrm{w}} = \mathcal{Z}_{\mathrm{w}^\star}$

$$\Psi_{\mathrm{w}}(\hat{Q},\hat{m},\hat{q}) \equiv \mathbb{E}_\xi\left[\mathcal{Z}_{\mathrm{w}^\star}\left(\hat{m}\hat{q}^{-1/2}\xi, \hat{m}\hat{q}^{-1}\hat{m}\right)\log\mathcal{Z}_{\mathrm{w}}\left(\hat{q}^{1/2}\xi, \hat{Q}+\hat{q}\right)\right]$$

- $\Psi_{\mathrm{out}}^{(r)}(Q) = \int \mathrm{d}y \int \mathrm{d}\tilde{\mathbf{z}}\, P_{\tilde{z}}(\tilde{\mathbf{z}};Q)P_{\mathrm{out}}(y \mid \tilde{\mathbf{z}})$

$$P_{\tilde{z}}(\tilde{\mathbf{z}};Q) = \frac{e^{-\frac{1}{2}\tilde{\mathbf{z}}^\intercal Q^{-1}\tilde{\mathbf{z}}}}{\det(2\pi Q)^{1/2}}$$

$$Q_{\mathrm{rs}}^{-1} = \begin{bmatrix} Q_{00}^{-1} & Q_{01}^{-1} & Q_{01}^{-1} & Q_{01}^{-1} \\ Q_{01}^{-1} & Q_{11}^{-1} & Q_{12}^{-1} & Q_{12}^{-1} \\ Q_{01}^{-1} & Q_{12}^{-1} & Q_{11}^{-1} & Q_{12}^{-1} \\ Q_{01}^{-1} & Q_{12}^{-1} & Q_{12}^{-1} & Q_{11}^{-1} \end{bmatrix}$$

$$\begin{aligned} Q_{00}^{-1} &= \left(Q^0 - rm(Q+(r-1)q)^{-1}m\right)^{-1} \\ Q_{01}^{-1} &= -\left(Q^0 - rm(Q+(r-1)q)^{-1}m\right)^{-1} m(q+(r-1)q)^{-1} \\ Q_{11}^{-1} &= (Q-q)^{-1} - (Q+(r-1)q)^{-1}q(Q-q)^{-1} \\ &\quad +(Q+(r-1)q)^{-1}m\left(Q^0 - rm(Q+(r-1)q)^{-1}m\right)^{-1} m(Q+(r-1)q)^{-1} \\ Q_{12}^{-1} &= -(Q+(r-1)q)^{-1}q(Q-q)^{-1} \\ &\quad +(Q+(r-1)q)^{-1}m\left(Q - rm(Q+(r-1)q)^{-1}m\right)^{-1} m(Q+(r-1)q)^{-1} \end{aligned}$$

$$\det Q_{\mathrm{rs}} = (Q-q)^{r-1}(Q+(r-1)q)\left(Q^0 - rm(Q+(r-1)q)^{-1}m\right)$$

Replica

$$\Psi_{\text{out}}^{(r)}(Q)\Big|_{\text{rs}} = \int \mathrm{d}y \int \mathrm{d}\tilde{\mathbf{z}} e^{-\frac{1}{2}\tilde{\mathbf{z}}^\top Q_{\text{rs}}^{-1}\tilde{\mathbf{z}} - \frac{1}{2}\log(\det(2\pi Q_{\text{rs}}))} P_{\text{out}}(y \mid \tilde{\mathbf{z}})$$

$$= \mathbb{E}_{y,\xi} \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi Q_{\text{rs}}))}$$

$$\times \int \mathrm{d}z^\star P_{\text{out}^\star}(y \mid z^\star)\, e^{-\frac{1}{2}Q_{00}^{-1}(z^\star)^2} \left[\int dz P_{\text{out}}(y \mid z) e^{-Q_{01}^{-1}z^\star z - \frac{1}{2}(Q_{11}^{-1} - Q_{12}^{-1})z^2 - Q_{12}^{-1/2}\xi z}\right]^r$$

similar operation

$$\Psi_{\text{out}}(Q, m, q; \rho_{\text{w}^\star}) \equiv \mathbb{E}_{y,\xi}\left[\mathcal{Z}_{\text{out}^\star}\left(y, mq^{-1/2}\xi, \rho_{\text{w}^\star} - mq^{-1}m\right)\log \mathcal{Z}_{\text{out}}\left(y, q^{1/2}\xi, Q - q\right)\right]$$

$$\mathcal{Z}_{\text{out}}(y, \omega, V) = \int dz \frac{1}{\sqrt{2\pi\Delta}} \exp\left(\frac{y - \text{sign}(z)}{2\Delta}\right) \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(z-\omega)^2}{2V}\right)$$

$$= \int_0^\infty dz \frac{1}{\sqrt{2\pi\Delta}} \exp\left(\frac{y-1}{2\Delta}\right) \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(z-\omega)^2}{2V}\right)$$

$$+ \int_{-\infty}^0 dz \frac{1}{\sqrt{2\pi\Delta}} \exp\left(\frac{y+1}{2\Delta}\right) \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(z-\omega)^2}{2V}\right)$$

$$= \mathcal{N}_y(1, \Delta^*)\left(1 + \text{erf}\left(\frac{\omega}{\sqrt{2V}}\right)\right) + \mathcal{N}_y(-1, \Delta^*)\frac{1}{2}\left(1 - \text{erf}\left(\frac{\omega}{\sqrt{2V}}\right)\right)$$

Finally, the replica symmetric solution of free energy is obtained as follows

$$\Phi_{\text{rs}}(\alpha) = \text{extr}_{Q,\hat{Q},q,\hat{q},m,\hat{m}}\left\{-m\hat{m} + \frac{1}{2}Q\hat{Q} + \frac{1}{2}q\hat{q} + \Psi_{\mathbf{w}}(\hat{Q}, \hat{m}, \hat{q}) + \alpha\Psi_{\text{out}}(Q, m, q; \rho_{\text{w}^\star})\right\}$$

Nishimori condition $Q = \rho_{\text{w}^\star}, \quad m = q = q, \quad \hat{Q} = 0, \quad \hat{m} = \hat{q} = \hat{q}$

$$\Phi(\alpha) = \text{extr}_{q,\hat{q}}\left\{-\frac{1}{2}q\hat{q} + \Psi_{\text{w}}(\hat{q}) + \alpha\Psi_{\text{out}}(q; \rho_{\text{w}^\star})\right\}$$

Replica

$$\hat{q} = -2\alpha\partial_q \Psi_{\text{out}}$$
$$q = -2\partial_{\hat{q}} \Psi_{\text{w}}$$

$$\Psi_{\text{w}}(\hat{q}) = \mathbb{E}_{\xi}\left[\mathcal{Z}_{\text{w}^\star}\left(\hat{q}^{1/2}\xi, \hat{q}\right)\log \mathcal{Z}_{\text{w}^\star}\left(\hat{q}^{1/2}\xi, \hat{q}\right)\right],$$
$$\Psi_{\text{out}}\left(q; \rho_{\text{w}^\star}\right) = \mathbb{E}_{y,\xi}\left[\mathcal{Z}_{\text{out}^\star}\left(y, q^{1/2}\xi, \rho_{\text{w}^\star} - q\right)\log \mathcal{Z}_{\text{out}^\star}\left(y, q^{1/2}\xi, \rho_{\text{w}^\star} - q\right)\right]$$
$$\mathcal{Z}_{\text{w}}(\gamma, \Lambda) \equiv \mathbb{E}_{w\sim P_{\text{w}}}\left[e^{-\frac{1}{2}\Lambda w^2 + \gamma w}\right] = \frac{e^{\frac{\gamma^2}{2(\Lambda+1)}}}{\sqrt{\Lambda+1}}$$
$$\mathcal{Z}_{\text{out}}(y, \omega, V) = \int dz \frac{1}{\sqrt{2\pi\Delta}} \exp\left(\frac{y - \text{sign}(z)}{2\Delta}\right) \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(z-\omega)^2}{2V}\right)$$
$$= \mathcal{N}_y\left(1, \Delta^*\right)\left(1 + \text{erf}\left(\frac{\omega}{\sqrt{2V}}\right)\right) + \mathcal{N}_y\left(-1, \Delta^*\right)\frac{1}{2}\left(1 - \text{erf}\left(\frac{\omega}{\sqrt{2V}}\right)\right)$$

define

$$f_{\text{w}^\star}(\gamma, \Lambda) \equiv \partial_\gamma \log \mathcal{Z}_{\text{w}^\star}(\gamma, \Lambda) = \frac{\gamma}{1+\Lambda}$$
$$f_{\text{out}^\star}(y, \omega, V) \equiv \partial_\omega \log \mathcal{Z}_{\text{out}^\star}(y, \omega, V) = \frac{\mathcal{N}_y\left(1, \Delta^\star\right) - \mathcal{N}_y\left(-1, \Delta^\star\right)}{\mathcal{Z}_{\text{out}^\star}(y, \omega, V)} \mathcal{N}_\omega(0, V)$$

equivalent to the definition in AMP

obtain

$$\hat{q} = \alpha \mathbb{E}_{y,\xi}\left[\mathcal{Z}_{\text{out}^\star}\left(y, q^{1/2}\xi, \rho_{\text{w}^\star} - q\right) f_{\text{out}^\star}\left(y, q^{1/2}\xi, \rho_{\text{w}^\star} - q\right)^2\right]$$
$$q = \mathbb{E}_{\xi}\left[\mathcal{Z}_{\text{w}^\star}\left(\hat{q}^{1/2}\xi, \hat{q}\right) f_{\text{w}^\star}\left(\hat{q}^{1/2}\xi, \hat{q}\right)^2\right]$$

Analytic solution

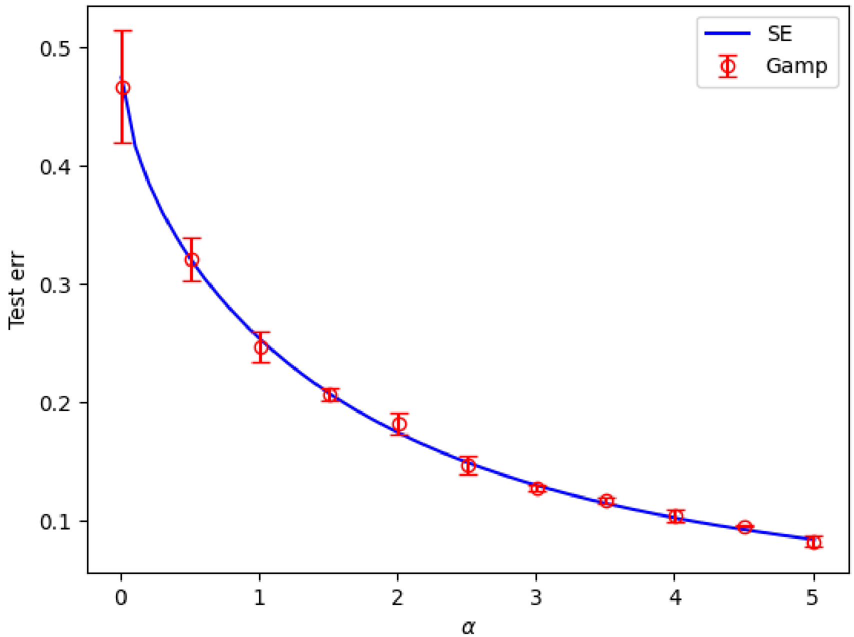
$$q = \frac{\hat{q}}{1+\hat{q}}$$

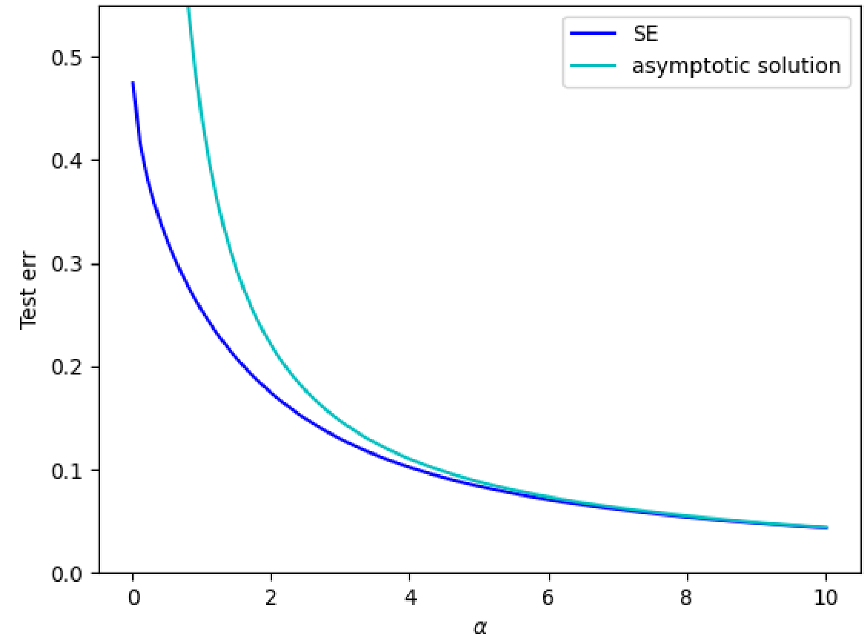
$$\hat{q} = \alpha \mathbb{E}_{y,\xi}\left[\mathcal{Z}_{\text{out}^\star}\left(y, q^{1/2}\xi, 1-q\right) f_{\text{out}^\star}\left(y, q^{1/2}\xi, 1-q\right)^2\right]$$
$$= 2\alpha \int D\xi y^2 \frac{\mathcal{N}_{\sqrt{q}\xi}\left(0, 1-q\right)^2}{\frac{1}{2}\left(1 + \text{erf}\left(\frac{\sqrt{q}\xi}{\sqrt{2(1-q)}}\right)\right)}$$
$$= \frac{2}{\pi}\frac{\alpha}{1-q}\int D\xi \frac{e^{-\frac{q_{\text{b}}\xi^2}{1-q}}}{\left(1 + \text{erf}\left(\frac{\sqrt{q}\xi}{\sqrt{2(1-q)}}\right)\right)}$$

same as SE equation

Conclusion

$$e_{\rm g}^{\rm bayes}(\alpha) = \frac{1}{\pi}\,\mathrm{acos}\left(\sqrt{q_{\rm b}}\right) \underset{\alpha\to\infty}{\simeq} \frac{1}{k\pi}\frac{1}{\alpha} \simeq \frac{0.4417}{\alpha}$$





Asymptotic solution $\alpha \to \infty,\ q \to 1$

$$\int D\xi \frac{e^{-\frac{q_{\rm b}\xi^2}{1-q_{\rm b}}}}{\left(1+\mathrm{erf}\left(\frac{\sqrt{q_{\rm b}}\xi}{\sqrt{2(1-q_{\rm b})}}\right)\right)} = \int \mathrm{d}\xi \frac{\frac{-e^{\frac{\xi^2(q_{\rm b}+1)}{2(1-q_{\rm b})}}}{\sqrt{2\pi}}}{\left(1+\mathrm{erf}\left(\frac{\sqrt{q_{\rm b}}\xi}{\sqrt{2(1-q_{\rm b})}}\right)\right)} \simeq \int \mathrm{d}\xi \frac{\frac{-e^{\frac{\xi^2}{1-q_{\rm b}}}}{\sqrt{2\pi}}}{\left(1+\mathrm{erf}\left(\frac{\xi}{\sqrt{2(1-q_{\rm b})}}\right)\right)}$$

$$= \frac{\sqrt{1-q_{\rm b}}}{\sqrt{2\pi}} \int \mathrm{d}\eta \frac{e^{-\eta^2}}{1+\mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)} = \frac{c_0}{\sqrt{2\pi}}\sqrt{1-q_{\rm b}}$$

$$q_{\rm b} = \frac{1}{2}\left(\alpha k\sqrt{\alpha^2 k^2+4} - \alpha^2 k^2\right) \underset{\alpha\to\infty}{\simeq} 1 - \frac{1}{\alpha^2 k^2}, \quad \hat{q}_{\rm b} = k^2\alpha^2$$

$$k \equiv \frac{2c_0}{\pi\sqrt{2\pi}} \simeq 0.720647$$