2023 年大创项目结题报告

神经网络的基本原理

具有连续权重的感知机的泛化误差

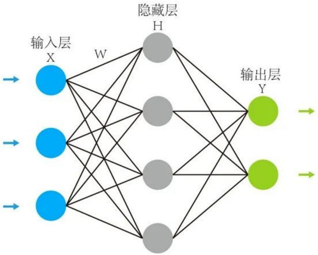
汇报人：李宇豪

项目成员：李宇豪、陈扬睿、李哲

指导老师：黄海平教授

2023 年 12 月 13 日

- **神经网络的工作模式**



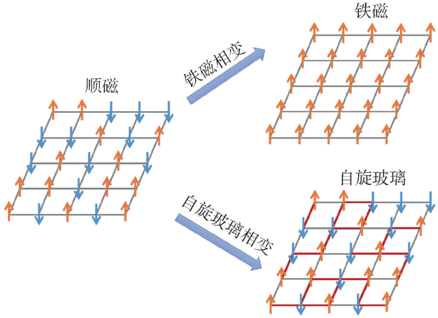
输出 $\hat{\boldsymbol{y}} = \sigma_2 \left[W_2 \cdot \sigma_1 \left(W_1 \boldsymbol{x} + b_1 \right) + b_2 \right]$

损失函数 $\mathcal{L} = \frac{1}{2} \left(\hat{\boldsymbol{y}} - \boldsymbol{y} \right)^2$

通过反向传播把 $\mathcal{L}$ 按梯度分配到不同层的权重

梯度下降法 **网络的学习**

- **自旋玻璃理论**



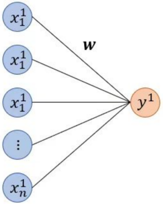
Sherrington—Kirkpatrick 模型

$$\mathcal{H} = -\sum_{i<j} J_{ij} \sigma_i \sigma_j$$

Giorgio Parisi 等人 复本方法、空腔方法

感知机的泛化误差

- **模型设定**



用固定权重 $\mathbf{w}^*$ 生成标签 $y = \text{sign}\left(\frac{1}{\sqrt{n}}\mathbf{X}\mathbf{w}^\star\right)$

感知机学习权重 $\mathbf{w} \to \mathbf{w}^*$

贝叶斯框架

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

- **泛化误差**

$$\varepsilon_{\text{gen}} = \lim_{n\to\infty}\mathbb{E}_{\mathbf{y},\mathbf{X}}\mathbb{1}[\mathbf{y} \neq \hat{y}(\hat{\mathbf{w}}(\alpha);\mathbf{x})] = \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\Theta(-\mathbf{z}\hat{\mathbf{z}})\right]$$

重参数化

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

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

泛化误差写为序参量的函数

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

其中

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

研究泛化误差随数据量密度的变化：

- 广义消息传递方程 $\to$ 状态演化方程
- 复本方法 $\to$ 复本对称解

广义消息传递方程

- **信念传播方程（BP Equation）**

  （空腔方法）

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

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

- **relax-BP Equation**

  （中心极限定理、泰勒展开）

$m_{i\to\mu}$ 的均值和方差

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

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

简化

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

其中

$$B^t_{\mu\to i} = X_{\mu i}\, f_{\text{out}}\left(\omega^t_{\mu\to i}, y_\mu, V^t_{\mu\to i}\right)$$

$$A^t_{\mu\to i} = -X^2_{\mu i}\, \partial_\omega f_{\text{out}}\left(\omega^t_{\mu\to i}, y_\mu, V^t_{\mu\to i}\right)$$

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

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

定义

$$\Sigma^{t+1}_{\mu\to i} = \frac{1}{\sum_\mu A^{t+1}_{\mu\to i}} \qquad R^{t+1}_{\mu\to i} = \frac{\sum_\mu B^{t+1}_{\mu\to i}}{\sum_\mu A^{t+1}_{\mu\to i}}$$

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

$\hat{w}$ 和 v 通过下式更新

$$\hat{w}^{t+1}_{\mu\to i} = f_{\text{w}}(\Sigma, R) = \frac{R^t_{\mu\to i}}{1+\Sigma^t_{\mu\to i}}$$

$$v^{t+1}_{\mu\to i} = \partial_R f_{\text{w}}(\Sigma, R) = \frac{1}{1+\Sigma^t_{\mu\to i}}$$

广义消息传递方程

- **广义消息传递方程（GAMP）算法总结**

初始化 $\hat{w}_i^0, v_i^0, f_{\text{out}}^0$

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

$$\omega_\mu^{t+1} = \frac{1}{\sqrt{n}}\sum_i X_{\mu i}\hat{w}_i^t - V_\mu^t f_{\text{out}}^t$$

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

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

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

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

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

- **状态演化方程（SE）**

定义

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

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

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

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

贝叶斯最优的框架有 Nishimori 条件 $q = m$

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

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

最终结果

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

复本方法

- **统计力学**

配分函数

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

自由能的淬火平均

$$\Phi = \frac{1}{n}\mathbb{E}_{\mathbf{y},\mathbf{X}}\ \log \mathcal{Z}(\mathbf{y},\boldsymbol{X}) \tag{1}$$

复本方法

$$\Phi = \frac{1}{n}\lim_{r\to 0}\frac{\partial \log \mathbb{E}_{\mathbf{y},\mathbf{X}}\ [\mathcal{Z}(\mathbf{y},\boldsymbol{X})^r]}{\partial r} \tag{2}$$

- **复本计算**（δ 函数的傅里叶变换、Laplace 近似）

$$\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathcal{Z}(\mathbf{y},\mathbf{X})^r\right] \propto \iint \mathrm{d}Q\,\mathrm{d}\hat{Q}\ e^{n\Phi^{(r)}(Q,\hat{Q})}$$

其中

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

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

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

鞍点方程

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

- **复本对称假设**

$$Q_{\mathrm{rs}} = \begin{pmatrix} Q^0 & m & \dots & m \\ m & Q & \dots & \dots \\ \dots & \dots & \dots & q \\ m & \dots & q & Q \end{pmatrix}$$

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

其中

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

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

自由能的复本对称解

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

其中,

$$\begin{aligned}
&\Psi_{\mathrm{w}}(\hat{Q},\hat{m},\hat{q})\equiv\mathbb{E}_{\xi}\left[\mathcal{Z}_{\mathrm{w}^{\star}}\left(\hat{m}\hat{q}^{-1/2}\xi,\hat{m}\hat{q}^{-1}\hat{m}\right)\log\mathcal{Z}_{\mathrm{w}}\left(\hat{q}^{1/2}\xi,\hat{Q}+\hat{q}\right)\right]\\
&\Psi_{\mathrm{out}}\left(Q,m,q;\rho_{\mathrm{w}^{\star}}\right)\equiv\mathbb{E}_{y,\xi}\left[\mathcal{Z}_{\mathrm{out}^{\star}}\left(y,mq^{-1/2}\xi,\rho_{\mathrm{w}^{\star}}-mq^{-1}m\right)\log\mathcal{Z}_{\mathrm{out}}\left(y,q^{1/2}\xi,Q-q\right)\right]\\
&\rho_{\mathbf{w}^{\star}}=\lim_{n\to\infty}\mathbb{E}_{\mathbf{w}^{\star}}\frac{1}{n}\left\|\mathbf{w}^{\star}\right\|_{2}^{2}
\end{aligned}\tag{3}$$

定义序参量

$$\hat{Q}=-2\alpha\partial_{Q}\Psi_{\mathrm{out}}\quad\hat{q}=-2\alpha\partial_{q}\Psi_{\mathrm{out}}\quad\hat{m}=\alpha\partial_{m}\Psi_{\mathrm{out}}$$

$$Q=-2\partial_{\hat{Q}}\Psi_{\mathrm{w}}\quad q=-2\partial_{\hat{q}}\Psi_{\mathrm{w}}\quad m=\partial_{\hat{m}}\Psi_{\mathrm{w}}$$

考虑到 Nishimori 条件，只关心

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

$$q=\mathbb{E}_{\xi}\left[\mathcal{Z}_{\mathrm{w}^{\star}}\left(\hat{q}^{1/2}\xi,\hat{q}\right)f_{\mathrm{w}^{\star}}\left(\hat{q}^{1/2}\xi,\hat{q}\right)^{2}\right]$$

复本对称计算

计算各个辅助函数的解析形式

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

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

$$\mathcal{Z}_{\text{w}^\star}(\gamma,\Lambda)=\frac{e^{\frac{\gamma^2}{2(\Lambda+1)}}}{\sqrt{\Lambda+1}}\quad f_{\text{w}^\star}(\gamma,\Lambda)=\frac{\gamma}{1+\Lambda}\quad \partial_\gamma f_{\text{w}^\star}(\gamma,\Lambda)=\frac{1}{1+\Lambda}$$

得出 q 和 $\hat{q}$ 的迭代形式

$$q=\frac{\hat{q}}{1+\hat{q}}\qquad \hat{q}=\frac{2}{\pi}\frac{\alpha}{1-q}\int D\xi\frac{e^{-\frac{q_{\text{b}}\xi^2}{1-q}}}{\left(1+\operatorname{erf}\left(\frac{\sqrt{q}\xi}{\sqrt{2(1-q)}}\right)\right)}$$

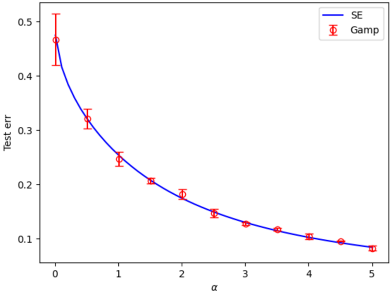
在 $\alpha\to\infty$ 时，有 $q\to 1$，在此极限下有

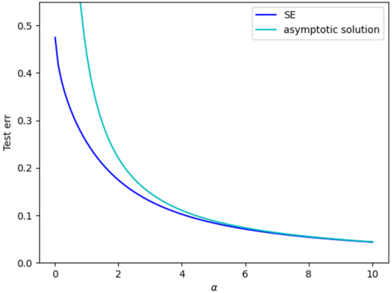
$$q_{\text{b}}=\frac{1}{2}\left(\alpha k\sqrt{\alpha^2k^2+4}-\alpha^2k^2\right)\underset{\alpha\to\infty}{\simeq}1-\frac{1}{\alpha^2k^2},\quad \hat{q}_{\text{b}}=k^2\alpha^2$$

得到泛化误差在大 α 时的渐进行为

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

- **GAMP 迭代求解与 SE、RS 的解析结果比较**





- **总结与展望**
  - 我们使用 GAMP、SE、Replica 等方法求解了感知机的泛化误差与数据量密度的关系；
  - 我们的创新点在于感知机的权重是连续的，而以前的工作主要研究离散权重的感知机（离散权重中存在相变）；
  - 这一类方法求解的是感知机（神经网络）收敛到稳态时的解，类似于统计力学中的平衡态；
  - 接下来的研究方向：
    - 更复杂的网络模型的稳态解（随机特征模型 RFM）
    - 感知机（神经网络）学习过程中的非平衡动力学

Thanks for Listening