

Phase Diagram of Deep Learning

《深度学习导论与理解》第七章

林元莘 洪以恒 杨博一

November 6, 2024

相图分析简介

相图是一种用来表示不同状态之间关系的图示。在传统物理学中，相图通过两个变量（如温度 T 和压力 P ）来决定物质的状态。例如水可以根据温度和压力处于固态、液态或气态。

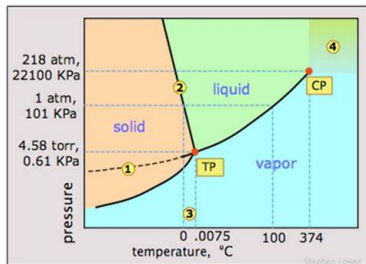


Figure: 水的相图

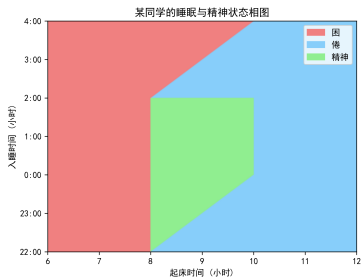


Figure: 某同学精神状态相图

神经网络的相图分析是一种研究神经网络在不同初始化条件下行为表现的方法，相图帮助我们理解模型参数如何影响网络的表现。神经网络的相图就是展示参数在不同初始化条件下行为表现的图表。

动机 - 隐式正则化与初始化

频率原则的局限性：未能提供定量方法表现神经网络参数学习过程，与难以区分神经网络与核方法，或不同神经网络之间的差异性。

- 用不同初始化大小（均值为0，不同标准差的高斯分布）的两层ReLU神经网络（宽度 $m = 1000$ ）学习四个数据点的结果。这里宽度 $m = 1000$ 是指隐藏层神经元的个数为1000。

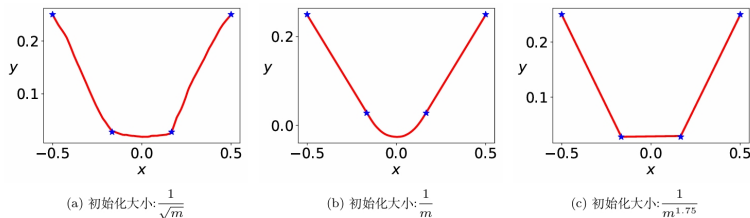


Figure: 不同初始化大小 \rightarrow 学到不同线性/非线性函数

不同初始化下的实验

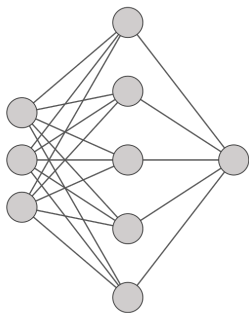


基本参数与分析

考虑一个具有两层结构和ReLU激活函数的神经网络，其数学表达式可以写为：

$$f_{\theta}(\mathbf{x}) = \sum_{k=1}^m \alpha_k \sigma(\omega_k^T \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d$$

$$\theta = \text{vec}(\theta_{\mathbf{a}}, \theta_{\mathbf{w}}), \quad \alpha_k^0 \sim \mathcal{N}(0, \beta_1^2), \quad \omega_k \sim \mathcal{N}(0, \beta_2^2 \mathbf{I}_d)$$



ReLU正齐次性: $\sigma(ax) = a\sigma(x), \quad \forall a > 0$

尺度放缩:

$$\bar{a}_k = \beta_1^{-1} a_k, \quad \bar{\omega}_k = \beta_2^{-1} \omega_k, \quad \bar{t} = \frac{1}{\beta_1 \beta_2} t$$

从动力学方程到缩放参数

在连续时间极限下, 我们有归一化的动力学方程:

$$\begin{aligned}\frac{d\bar{a}_k}{d\bar{t}} &= -\frac{1}{n\kappa'} \sum_{i=1}^n \kappa \sigma(\bar{w}_k^T x_i) \left(\kappa \sum_{k'=1}^m \bar{a}_{k'} \sigma(\bar{w}_{k'}^T x_i) - y_i \right), \\ \frac{d\bar{w}_k}{d\bar{t}} &= -\frac{\kappa'}{n} \sum_{i=1}^n \kappa \bar{a}_k \sigma'(\bar{w}_k^T x_i) x_i \left(\kappa \sum_{k'=1}^m \bar{a}_{k'} \sigma(\bar{w}_{k'}^T x_i) - y_i \right),\end{aligned}$$

其中 κ 和 κ' 分别为能量缩放参数和动力学缩放参数. 再定义

$$\gamma = -\lim_{m \rightarrow \infty} \frac{\log \kappa}{\log m}, \quad \gamma' = -\lim_{m \rightarrow \infty} \frac{\log \kappa'}{\log m}.$$

通过相对距离(RD)识别三种区域

相对距离(Relative Distance):

$$\text{RD}(\theta_w(t)) = \frac{\|\theta_w(t) - \theta_w(0)\|_2}{\|\theta_w(0)\|_2}.$$

当 $m \rightarrow \infty$ 时,

线性区域: $\sup_{t \in [0, +\infty)} \text{RD}(\theta_w(t)) \rightarrow 0,$

临界区域: $\sup_{t \in [0, +\infty)} \text{RD}(\theta_w(t)) = \mathcal{O}(1),$

凝聚区域: $\sup_{t \in [0, +\infty)} \text{RD}(\theta_w(t)) \rightarrow \infty.$

线性与非线性行为的界定

对任意训练时刻 t , 如果参数 θ 变化量足够小那么模型 $f_{\theta}(\mathbf{x})$ 可以被以下模型近似:

$$f_{\theta}(\mathbf{x}) \approx f_{\theta(0)}(\mathbf{x}) + \nabla_{\theta} f_{\theta(0)}(\mathbf{x}) \cdot (\theta(t) - \theta(0))$$

$\theta_{\mathbf{w}}(t)$ 要保持在 $\theta_{\mathbf{w}}(0)$ 的一个小邻域内有效。我们将这个线性近似有效的区域称为“线性区域”

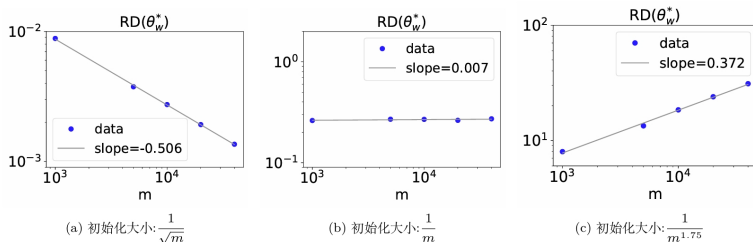


Figure: $RD(\theta_w^*)$ 与 m 的关系, when $m \rightarrow \infty$

$RD(\theta_w^*)$ 实验

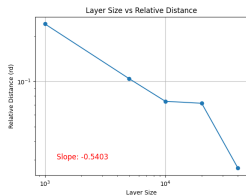


Figure: 初始化大小 $\frac{1}{\sqrt{m}}$

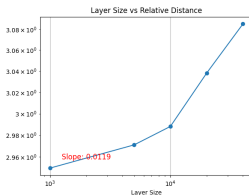


Figure: 初始化大小 $\frac{1}{m}$

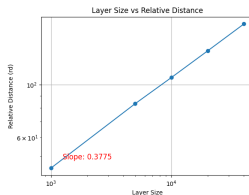
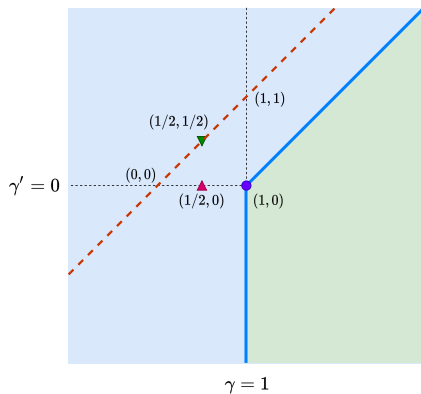


Figure: 初始化大小 $\frac{1}{m^{1.75}}$

Phase Diagram



Linear regime

Condensed regime

Critical regime

Examples:

● Xavier, Mean field

▲ NTK

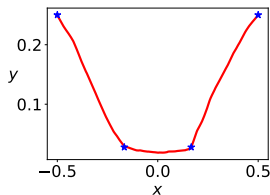
- - E at el. (2020)

▼ LeCun, He

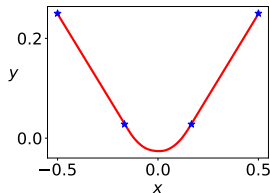
$$a_k^0 \sim N(0, \beta_1^2), \quad \mathbf{w}_k^0 \sim N(0, \beta_2^2 \mathbf{I}_d)$$

$$\gamma = \lim_{m \rightarrow \infty} -\frac{\log \beta_1 \beta_2 / \alpha}{\log m}, \quad \gamma' = \lim_{m \rightarrow \infty} -\frac{\log \beta_1 / \beta_2}{\log m}$$

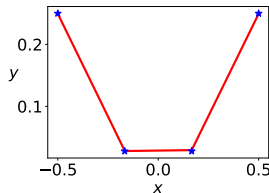
相图对应的三种区域



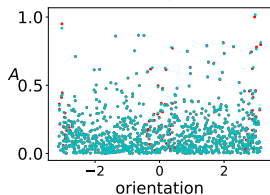
合成数据, $\gamma = 0.5$



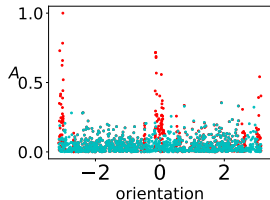
合成数据, $\gamma = 1.0$



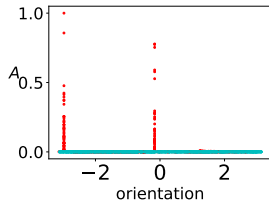
合成数据, $\gamma = 1.75$



MNIST, $\gamma = 0.5$

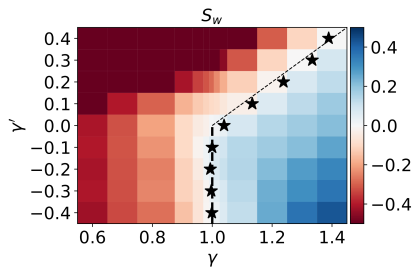


MNIST, $\gamma = 1.0$

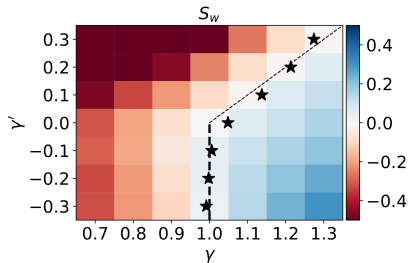


MNIST, $\gamma = 1.75$

实验相图

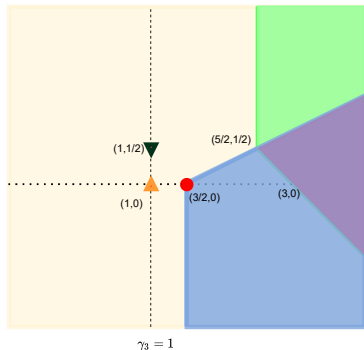


合成数据



MNIST

三层网络的理论相图



Example :

● Xavier

▲ NTK

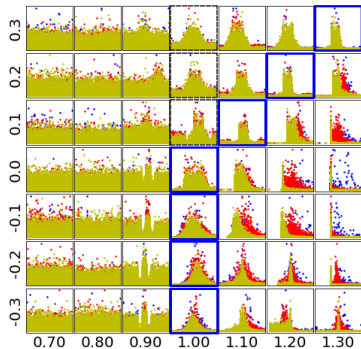
▼ Lecun, He

$$\mathbf{a}_k \sim \mathcal{N}(0, \beta_3^2), \mathbf{W}_{ij}^{[2]} \sim (0, \beta_2^2), \mathbf{W}_{ij}^{[1]} \sim (0, \beta_1^2)$$

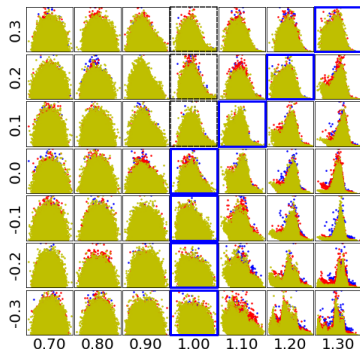
$$\gamma_3 = \lim_{m \rightarrow \infty} -\frac{\log \beta_1 \beta_2 \beta_3 / \alpha}{\log m}, \quad \gamma_2 = \lim_{m \rightarrow \infty} -\frac{\log \beta_3 / \beta_1}{\log m}$$

CR is short for Condensed Regime, LR is short for Linear Regime

相图上不同位置的特征分布



合成数据



MNIST

- **本章总结:** 神经网络的相图分析是一种研究神经网络在不同初始化条件下行为表现的方法, 相图帮助我们理解模型参数如何影响网络的表现。
- **相图分析的局限性** 有限宽度神经网络在实际应用中存在偏差;改进对有限宽长度的分析方法; 可引入正则化技术和扩展到更复杂的模型。
- **未来展望:** 对于**凝聚现象**的具体分析以及其它复杂神经网络(全连接神经网络, CNN, RNN 等等)的凝聚现象探讨深度学习导论与理论第八章内容

谢谢！