CS 839: Theoretical Foundations of Deep Learning **Spring 2022**

Lecture 16 Mean Field Analysis of Neural Networks III

Instructor: Yingyu Liang *Date: March* 24^{th}, 2022 *Scriber: Haotian Shi*

1 Assumptions

This section presents several assumptions needed for theoretical analysis.

Assumption 1:(activation function)

Assume that $\sigma(\theta, x)$ satisfies the following condition:

$$\forall\theta,\ \mathbb{E}_x[\sigma(\theta, x)^2] \leq B_r^2$$

Assumption 2:(properties of the loss function). We assume:

1. $l(\widehat{y}, y)$ is convex on $\widehat{y}$.

2. $l(\widehat{y}, y)$ is bounded below, i.e., $l(\widehat{y}, y) \geq B_l$.

3. $l(\widehat{y}, y)$ is $L_1 - Lipschitz$ and has $L_2 - Lipschitz$ continous gradient. i.e., $|l'(\widehat{y}, y)| \leq L_1; |l'(\widehat{y}_1, y) - l'(\widehat{y}_2, y)| \leq L_2|\widehat{y}_1 - \widehat{y}_2|$.

Assumption 3:(properties of the feature activation function $h^{'}$). Under *Assumption* 1, we further assume:

1. for all x, $\sigma(\theta, x)$ is second-order differentiable on θ.

2. for all x and θ, we assume $|\sigma(\theta, x)| \leq C_1\|\theta\| + C_2$; $\|\nabla_\theta\sigma(\theta, x)\| \leq C_3$; $\|\nabla_\theta^2\sigma(\theta, x)\| \leq C_3$.

As for the smoothness conditions in Assumptions 3, they hold for many feature functions, e.g. tanh, sigmoid, smoothed relu.

Assumption 4:(initial value). We assume: $Q'(p_0) \leq \infty$.

Assumption 4 holds for common distributions that have bounded second moments and are absolutely continuous with respect to the Lebesgue measure. A safe setting of $p_0$might be a standard Gaussian distribution.

2 Covergence of GD

It is not hard to observe that the continuous NN learning is a convex optimization problem in the infinite dimensional measure space. So by exploiting the convexity, we describe the properties for the solution of $Q'(p)$ as follows.

Proposition (Global Optimal Solution) Suppose Assumption 2 and 3 hold, $Q'(p)$ is convex with respect to p and has a unique optimal solution p^*, a.e., which satisfies:

$$p^* = \frac{\exp\left(-\frac{\lambda_1}{2\lambda_3}|u|^2 - \frac{\lambda_2}{2\lambda_3}\|\theta\|^2 - \frac{u}{\lambda_3}E_{(x,y)}\left[l'(f(\omega^*, \rho^*, x), y)\sigma(\theta, x)\right]\right)}{C_5} = \frac{\exp(-\frac{\psi_{p^*}}{\lambda_3})}{C_5}$$

where $C_5$is a finite constant for normalization. Moreover, we have $p^* > 0$. Therefore, we can get that:

$$Q'(p) = E_{(x,y)}\left[l(\int \sigma(\theta, x)p(\theta, u)dud\theta, y)\right] + \int \left(\frac{\lambda_1}{2}|u|^2 + \frac{\lambda_2}{2}\|\theta\|^2\right) pdud\theta + \lambda_3 \int p \ln pd\theta du$$

Theorem (Convergence of NGD) Uner Assumption 2, 3, and 4, and suppose that p_t evolves, then $p_t$coverges weakly to p_*. Moreover,

$$\lim_{t\to\infty} Q(p_t) = Q(p^*)$$

Proof of sketch:

In this proof, we use θ to denote $[\theta, u]$.

Step 1. We prove that $E_{p_t}\|\theta\|^2 \leq B_M, \forall t \geq 0$, where B_M is a finite constant.

Step 2. From Step 1, the second moment of $p_t(\theta')$ is uniformly bounded by B_M. So $p_t(\theta')$ is uniformly tight. Thus there exsits a $p_\infty$and a subsequence p_k with $k \to \infty$,$p_k$converges weakly to p_∞. Let:

$$\psi_p(\theta, u) = \frac{\lambda_1}{2}|u|^2 + \frac{\lambda_2}{2}\|\theta\|^2 + uE_{(x,y)}[l'(f_p(x), y)\sigma(\theta, x)]$$

We prove:

$$\lim_{k\to\infty} \int \|\nabla\psi_{p_k} - \nabla\psi_{p_\infty}\|^2 p_k d\widetilde{\theta} = 0$$

Step 3. We further prove:

$$\lim_{k\to\infty} \int |p_k^{1/2} \exp(\frac{\psi_{p_\infty}}{2\lambda_3}) - c_k|^2 G(\widetilde{\theta})d\theta = 0,$$

where

$$G(\widetilde{\theta}) \propto \exp(-\frac{\lambda_1}{2\lambda_3}|u|^2 - \frac{\lambda_2}{2\lambda_3}\|\theta\|^2)$$

Step 4. Because c_k is bounded, we can take a sub-sequence $t_k$with $\lim_{k\to\infty} c_{t_k} = c_\infty$. Then:

$$\lim_{k\to\infty} \int |p_k^{1/2} \exp(\psi_{p_\infty}/2\lambda_3) - c_\infty|^2 G(\widetilde{\theta})d\widetilde{\theta} = 0$$

Furthermore, there exists a sub-sequence$\tau_k \subseteq t_k$ such that:

$$\lim_{k\to\infty} p_{\tau_k} \exp(\psi_{p_\infty}/2\lambda_3) = c_\infty, a.e.$$

It follows that:

$$p_{\tau_k} \to c_\infty^2 \exp(\psi_{p_\infty}/\lambda_3) = \widetilde{p}_\infty, a.e.$$

Let $\widetilde{p}_\infty = c_\infty^2(-\psi_{p_\infty}/\lambda_3)$. We prove $p_\infty = \widetilde{p}_\infty, a.e.$

Step 5. Finally, we prove that $\widetilde{p}_\infty = p_\infty = p^*$. a.e. and $\lim_{t\to\infty} Q(p_t) = Q(P_*)$.