Asymptotic Properties for Bayesian Neural Network in Besov Space - Theorem 2

이경원

February 1, 2023

정리 2

아래의 정리 은 Polson and Ročková [2018]의 정리 6.2를 베소프 공간으로 확장한 것입니다. 해당 정리는 베이즈 ReLU 신경망이 spike-and-slab 사전분포 하에서 평활 모수를 모르는 경우에도 최적의 베이즈 추론이 가능함을 의미합니다.

정리 (Theorem 2 in Lee and Lee [2022]). *Assume model*

$$y_i = f_0(X_i) + \xi_i, \quad \xi_i \overset{iid}{\sim} N(0, \sigma^2) \quad i = 1, \cdots, n,$$

Let

$$\tilde{L}_n(H) = \lceil H(\log n) \rceil \vee 1,\ \tilde{W}_n(H, N) = HN,\ \tilde{S}_n(H, N) = HN\tilde{L}_n(H),\ \tilde{B}_n(H, N) = N^H. \tag{1}$$

Consider the following prior

$$N \overset{d}{\equiv} 1 \vee \lceil Z/(\log n)^2 \rceil, \quad \pi_Z(Z) = \frac{\lambda_N^Z}{Z!(e^{\lambda_N} - 1)} \quad for\ Z = 1, 2, \cdots, \tag{2}$$

and prior distribution

$$\pi(\theta_j|\gamma_j, L, W, S, B) = \gamma_j \tilde{\pi}(\theta_j|L, W, S, B) + (1 - \gamma_j)\delta_0(\theta_j),$$
$$\pi(\gamma|L, W, S, B) = 1/\binom{T}{S},$$
$$\pi(L = L_n) = \pi(W = W_n) = \pi(S = S_n) = \pi(B = B_n) = 1.$$

of θ given $\tilde{L}_n(H_n),\ \tilde{W}_n(H_n, N),\ \tilde{S}_n(H_n, N),\ \tilde{B}_n(H_n, N)$ on the function space

$$\Phi = \bigcup_{n=1}^{\infty} \bigcup_{N=1}^{\infty} \Phi(\tilde{L}_n(H_n),\ \tilde{W}_n(H_n, N),\ \tilde{S}_n(H_n, N),\ \tilde{B}_n(H_n, N)) \tag{3}$$

for any $H_n \to \infty$ *and* $\lambda_N > 0$. *Suppose that* $0 < F < \infty$, $0 < p, q \leq \infty$ *and* $d(1/p - 1/2)_+ < s < \min\{m, m-1+1/p\}$. *The posterior distribution concentrates at the true function with a rate* $\epsilon_n = n^{-s/(2s+d)}(\log n)^{3/2}$. *That is,*

$$\Pi\left(f_\theta \in \Phi \cap \mathcal{UB}(F) : \|f_\theta - f_0\|_n > M_n \epsilon_n | \mathbb{D}_n\right) \to 0$$

in $P_{f_0}^{(n)}$*-probability as* $n \to \infty$ *for any* $M_n \to \infty$.

Proof. Let $\mathcal{F} = \Phi \cap \mathcal{UB}(F)$. From **Lemma 3**, it is enough to show that there exist a constant $C' > 2/\sigma^2$ and $\mathcal{F}_n \subset \mathcal{F}$ which satisfy

(a) $\sup_{\epsilon > \epsilon_n} \log N(\epsilon/36, A_{\epsilon,1} \cap \mathcal{F}_n, \|\cdot\|_n) \leq n\epsilon_n^2$

(b) $-\log \Pi(A_{\epsilon_n,1}) \leq C' n \epsilon_n^2$

(c) $\Pi(\mathcal{F} - \mathcal{F}_n) = o\left(e^{-(C'\sigma^2+2)n\epsilon_n^2}\right)$

for sufficiently large n. From

$$L_n = O(\log n),\ W_n = O(N_n),\ S_n = O(N_n \log n), \tag{4}$$

we can choose

$$H_0 > \sup_n \{L_n/\log n,\ W_n/N_n,\ S_n/(N_n \log n),\ \Xi\}. \tag{5}$$

Let $\tilde{N}_n = C_N N_n$,

$$\mathcal{F}_n = \mathcal{UB}(F) \cap \left(\bigcup_{N=1}^{\tilde{N}_n} \Phi(\tilde{L}_n(H_0),\ \tilde{W}_n(H_0, N),\ \tilde{S}_n(H_0, N),\ \tilde{B}_n(H_0, N)) \right)$$

for sufficiently large $C_N > 0$ and $\pi_N(N)$ be a density function of N. First, show that (a) holds. From **Lemma 4**,

$$\begin{aligned} & N\left(\frac{\epsilon_n}{36}, \mathcal{F}_n, \|\cdot\|_{L^\infty}\right) \\ & \leq \sum_{N=1}^{\tilde{N}_n} \left(\frac{72}{\epsilon_n} \tilde{L}_n(H_0) (\tilde{B}_n(H_0, N) \vee 1)^{\tilde{L}_n(H_0)} (\tilde{W}_n(H_0, N) + 1)^{2\tilde{L}_n(H_0)} \right)^{\tilde{S}_n(H,N)+1} \\ & \leq \tilde{N}_n \left(\frac{72}{\epsilon_n} \tilde{L}_n(H_0) (\tilde{B}_n(H_0, \tilde{N}_n) \vee 1)^{\tilde{L}_n(H_0)} (\tilde{W}_n(H_0, \tilde{N}_n) + 1)^{2\tilde{L}_n(H_0)} \right)^{\tilde{S}_n(H_0,\tilde{N}_n)+1} \end{aligned} \tag{6}$$

and

$$
\begin{aligned}
&\sup_{\epsilon>\epsilon_n} \log N\left(\frac{\epsilon}{36}, A_{\epsilon,1}\cap\mathcal{F}_n, \|\cdot\|_n\right) \\
&\leq \sup_{\epsilon>\epsilon_n} \log N\left(\frac{\epsilon}{36}, A_{\epsilon,1}\cap\mathcal{F}_n, \|\cdot\|_{L^\infty}\right) \\
&\leq \sup_{\epsilon>\epsilon_n} \log N\left(\frac{\epsilon}{36}, \mathcal{F}_n, \|\cdot\|_{L^\infty}\right) \\
&\leq \log N\left(\frac{\epsilon_n}{36}, \mathcal{F}_n, \|\cdot\|_{L^\infty}\right) \\
&\leq \log \tilde{N}_n \\
&\quad + \left[\tilde{S}_n(H_0,\tilde{N}_n)+1\right] \log\left(\frac{72}{\epsilon_n}\tilde{L}_n(H_0)(\tilde{B}_n(H_0,\tilde{N}_n)\vee 1)^{\tilde{L}_n(H_0)}(\tilde{W}_n(H_0,\tilde{N}_n)+1)^{2\tilde{L}_n(H_0)}\right) \\
&\lesssim \tilde{N}_n(\log n)^3 \\
&\lesssim n\epsilon_n^2.
\end{aligned}
\tag{7}
$$

for sufficiently large n. Next, show that (b) holds. Note $N_n(\log n)^3 \lesssim n\epsilon_n^2$ and

$$
L_n \leq \tilde{L}_n(H_n),\ W_n \leq \tilde{W}_n(H_n,N_n),\ S_n \leq \tilde{S}_n(H_n,N_n),\ B_n \leq \tilde{B}_n(H_n,\tilde{N}_n) \tag{8}
$$

for $N_n,\ L_n,\ W_n,\ S_n,\ B_n$ in **Theorem 1** and sufficiently large n. Thus, there exists a constant $D>0$ such that

$$
\pi_N(N_n) \gtrsim \exp\left(-N_n(\log n)^2 \log\frac{N_n}{\lambda_N}\right) \gtrsim \exp\left(-Dn\epsilon_n^2\right) \tag{9}
$$

and

$$
\begin{aligned}
&\Pi(f_\theta\in\mathcal{F}_n : \|f-f_0\|_n \leq \epsilon_n) \\
&\geq \pi_N(N_n)\Pi(f_\theta\in\Phi(L_n,W_n,S_n,B_n) : \|f_\theta-f_0\|_n \leq \epsilon_n | N_n) \\
&\gtrsim \exp\left(-(C+D)n\epsilon_n^2\right)
\end{aligned}
\tag{10}
$$

holds for sufficiently large n. (b) holds for $C' = \max\left\{C+D, 1+2/\sigma^2\right\}$. From

$$
\Pi(\mathcal{F}-\mathcal{F}_n) \leq \pi_N(N>\tilde{N}_n)
$$

and *Chernoff bound*, for any positive number $t,\ Z_0>0$,

$$
P(Z>Z_0) < e^{-t(Z_0+1)}\mathbb{E}\left[e^{tZ}\right] \lesssim e^{-t(Z_0+1)}\left(\exp\left(e^t\lambda_N\right)-1\right). \tag{11}
$$

Letting $t=\log Z_0$, we get

$$
P(Z>Z_0) \lesssim e^{-(Z_0+1)\log Z_0}\left(\exp(Z_0\lambda_N)-1\right). \tag{12}
$$

Thus,

$$
\begin{gathered}
\pi_N(N>\tilde{N}_n) \lesssim e^{-\left[(\tilde{N}_n+1)\log\tilde{N}_n+\tilde{N}_n\lambda_N\right](\log n)^2}, \\
(C'\sigma^2+2)n\epsilon_n^2 + \lambda_N\tilde{N}_n(\log n)^2 - (\tilde{N}_n+1)\log\tilde{N}_n(\log n)^2 \to -\infty
\end{gathered}
\tag{13}
$$

for sufficiently large $C_N>0$. (c) holds.

참고문헌

K. Lee and J. Lee. Asymptotic properties for bayesian neural network in besov space. In *Advances in Neural Information Processing Systems*, 2022.

N. G. Polson and V. Ročková. Posterior concentration for sparse deep learning. *Advances in Neural Information Processing Systems*, 31, 2018.