Frequentist coverage guarantees for Empirical Bayesian deep neural networks

S.E.M.P. Franssen,[@*] B.T. Szabó[@†]

[@]s.e.m.p.franssen@tudelft.nl, [@]botond.szabo@unibocconi.it, [*]Delft Institute of Applied Mathematics, TU Delft, [†]Department of Decision sciences, Bocconi University

Introduction

We provide the first methodology for uncertainty quantification using deep neural networks with theoretical guarantees.

Selling points

- First methodology with theoretical guarantees
- Significantly faster than bootstrap and fully Bayesian approaches
- Easy to implement
- Easily adapted to new statistical tasks

Set up

$$
\begin{tikzcd}
 & \text{Data 1st half} \arrow[r] & \text{DNN} \arrow[r] & \{\hat{\phi}_j\}_{j=1}^{K} \arrow[d] \\
\text{Data} \arrow[ur] \arrow[dr] & & \text{Prior : } w_j \sim \Pi \arrow[r] & \sum_{j=1}^{K} w_j \hat{\phi}_j \arrow[d] \\
 & \text{Data 2nd half} \arrow[rr] & & \text{Posterior } \Pi\left(..|\text{Data}\right)
\end{tikzcd}
$$

Earlier results on convergence rates

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Assumptions

- True function is β-smooth
- Have $k_n = n^{1/(2\beta+d)}$ basis functions
- Have found a good (local) optimizer of the deep neural network
- Sparse deep neural networks*
- Near orthogonal basis functions*

* Needed in theory to get the best bounds

Simulation studies

Regression

Classification

Theoretical guarantee for regression

Denote $\epsilon_n = n^{\frac{-\beta}{2\beta+d}}$ and $\tilde{\epsilon_n} = \epsilon_n \log(n)^3$.

Theorem 1. *Let $\beta, M > 0$. Under some conditions*

- *The posterior contracts at near minimax rate:*
$$\limsup_{n\to\infty} \sup_{f_0 \in W_M^\beta} \mathbb{E}_{f_0}\left(\Pi\left(\|f - f_0\|_2 \geq M_n \tilde{\epsilon}_n | \mathcal{D}_n\right)\right) = 0,$$
for all $M_n \to \infty$.
- *The credible balls have uniform near optimal coverage: There exists $L_{\epsilon,\alpha}$ such that if $B(c_\alpha, R_\alpha)$ is an α-credible ball, the ball $B = B(c_\alpha, L_{\epsilon,\alpha} \log(n)^3 R_\alpha)$ satisfies*
$$\liminf_n \inf_{f_0 \in W_M^\beta([0,1]^d)} \mathbb{P}_{f_0}^{(n)}(f_0 \in B) \geq 1 - \epsilon$$
- *The credible balls have uniform near optimal size:*
$$\liminf_n \inf_{f_0 \in W_M^\beta([0,1]^d)} \mathbb{P}_{f_0}^{(n)}(R_\alpha \leq C\epsilon_n) \geq 1 - \epsilon$$
for some large enough $C > 0$.

Conclusion

Empirical Bayesian deep neural networks provide a great way to do uncertainty quantification.