

Corrections to ‘‘Generalization Bounds via Information Density and Conditional Information Density’’

Fredrik Hellström, *Student Member, IEEE*, Giuseppe Durisi, *Senior Member, IEEE*

Abstract—An error in the proof of the data-dependent tail bounds on the generalization error presented in Hellström and Durisi (2020) is identified, and a correction is proposed. Furthermore, we note that the absolute continuity requirements in Hellström and Durisi (2020) need to be strengthened to avoid measurability issues.

I. DATA-DEPENDENT BOUNDS IN [1, EQS. (26), (34), (95), AND (98)]

In the proof of [1, Eq. (26)], we incorrectly claimed that [1, Eq. (32)] implies [1, Eq. (26)]. The issue is that [1, Eq. (32)] holds for a *fixed* λ , whereas, for [1, Eq. (26)] to hold, [1, Eq. (32)] needs to hold uniformly over all $\lambda \in \mathbb{R}$.

This issue can be fixed as follows. Since $\text{gen}(w, \mathbf{Z})$ is σ/\sqrt{n} -sub-Gaussian with zero mean under $P_{\mathbf{Z}}$ for all w , we can apply [2, Thm. 2.6.(IV)] (with $\lambda = 1 - 1/n$ therein) to conclude that

$$\mathbb{E}_{P_{\mathbf{Z}}} \left[\exp \left(\frac{n-1}{2\sigma^2} (\text{gen}(w, \mathbf{Z}))^2 \right) \right] \leq \sqrt{n}. \quad (1)$$

Taking the expectation with respect to P_W , changing measure to $P_{W\mathbf{Z}}$, and rearranging terms, we obtain

$$\mathbb{E}_{P_{W\mathbf{Z}}} \left[\exp \left(\frac{n-1}{2\sigma^2} (\text{gen}(W, \mathbf{Z}))^2 - \log \sqrt{n} - \iota(W, \mathbf{Z}) \right) \right] \leq 1. \quad (2)$$

Proceeding as in [1, Cor. 2], with an additional use of Jensen’s inequality, we find that with probability at least $1 - \delta$ under $P_{\mathbf{Z}}$,

$$\begin{aligned} & \left| \mathbb{E}_{P_{W|\mathbf{Z}}} [\text{gen}(W, \mathbf{Z})] \right| \\ & \leq \sqrt{\frac{2\sigma^2}{n-1} \left(D(P_{W|\mathbf{Z}} \| P_W) + \log \frac{\sqrt{n}}{\delta} \right)}. \end{aligned} \quad (3)$$

Similarly, proceeding as in the proof of [1, Eq. (34)], we find that with probability at least $1 - \delta$ under $P_{W\tilde{\mathbf{Z}}\mathbf{S}}$,

$$|\text{gen}(W, \mathbf{Z})| \leq \sqrt{\frac{2\sigma^2}{n-1} \left(\iota(W, \mathbf{Z}) + \log \frac{\sqrt{n}}{\delta} \right)}. \quad (4)$$

The issue reported in this note also affects the data-dependent tail bounds for the random-subset setting reported in [1, Eqs. (95) and (98)]. To fix it, we use that for any fixed $(w, \tilde{\mathbf{z}})$, the random variable $\widehat{\text{gen}}(w, \tilde{\mathbf{z}}, \mathbf{S})$ is $1/\sqrt{n}$ -sub-Gaussian with zero mean

under $P_{\mathbf{S}}$. Applying [2, Thm. 2.6.(IV)] with $\lambda = 1 - 1/n$ we obtain

$$\mathbb{E}_{P_{\mathbf{S}}} \left[\exp \left(\frac{n-1}{2} (\widehat{\text{gen}}(w, \tilde{\mathbf{z}}, \mathbf{S}))^2 \right) \right] \leq \sqrt{n}. \quad (5)$$

Taking the expectation with respect to $P_{W\tilde{\mathbf{Z}}}$, changing measure to $P_{W\tilde{\mathbf{Z}}\mathbf{S}}$, and rearranging terms, we conclude that

$$\begin{aligned} & \mathbb{E}_{P_{W\tilde{\mathbf{Z}}\mathbf{S}}} \left[\exp \left(\frac{n-1}{2} (\widehat{\text{gen}}(W, \tilde{\mathbf{Z}}, \mathbf{S}))^2 \right. \right. \\ & \quad \left. \left. - \log \sqrt{n} - \iota(W, \mathbf{S}|\tilde{\mathbf{Z}}) \right) \right] \leq 1. \end{aligned} \quad (6)$$

Proceeding as in [1, Cor. 6], we conclude that with probability at least $1 - \delta$ under $P_{\tilde{\mathbf{Z}}\mathbf{S}}$,

$$\begin{aligned} & \mathbb{E}_{P_{W|\tilde{\mathbf{Z}}\mathbf{S}}} [\widehat{\text{gen}}(W, \tilde{\mathbf{Z}}, \mathbf{S})] \\ & \leq \sqrt{\frac{2}{n-1} \left(D(P_{W|\tilde{\mathbf{Z}}\mathbf{S}} \| P_{W|\tilde{\mathbf{Z}}}) + \log \frac{\sqrt{n}}{\delta} \right)}. \end{aligned} \quad (7)$$

Furthermore, with probability at least $1 - \delta$ under $P_{W\tilde{\mathbf{Z}}\mathbf{S}}$,

$$\left| \widehat{\text{gen}}(W, \tilde{\mathbf{Z}}, \mathbf{S}) \right| \leq \sqrt{\frac{2}{n-1} \left(\iota(W, \mathbf{S}|\tilde{\mathbf{Z}}) + \log \frac{\sqrt{n}}{\delta} \right)}. \quad (8)$$

To summarize, the data-dependent tail bounds reported in [1, Eqs. (26), (34), (95), and (98)] should be replaced with (3), (4), (7), and (8) respectively.

Note that the data-independent tail bounds that we provide in [1, Eqs. (27), (35), (41), (42), (96), (99), (101), and (102)] still hold verbatim, although their proofs need to be modified. Specifically, for a fixed λ , one needs to first replace the information measure appearing in the bounds with its data-independent relaxation. The desired bounds then follow by setting λ equal to a suitably chosen, data-independent constant. Consider for example the data-independent bound in [1, Eq. (27)]. To obtain it, we first use [1, Eq. (33)] in [1, Eq. (32)], which results in

$$\begin{aligned} & P_{\mathbf{Z}} \left[\frac{\lambda^2 \sigma^2}{2n} - \lambda \mathbb{E}_{P_{W|\mathbf{Z}}} [\text{gen}(W, \mathbf{Z})] + \frac{\mathbb{E}_{P_{\mathbf{Z}}}^{1/t} [D(P_{W|\mathbf{Z}} \| P_W)^t]}{\delta^{1/t}} \right. \\ & \quad \left. + \log \frac{1}{\delta} \geq 0 \right] \geq 1 - 2\delta. \end{aligned} \quad (9)$$

The desired result follows by setting $\lambda = \pm \sqrt{\frac{a}{b}}$, where $a = \mathbb{E}_{P_{\mathbf{Z}}}^{1/t} [D(P_{W|\mathbf{Z}} \| P_W)^t] / \delta^{1/t} + \log \frac{1}{\delta}$ and $b = \sigma^2 / (2n)$, and then replacing δ with $\delta/2$.

F. Hellström and G. Durisi are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden, (e-mail: {fhells,durisi}@chalmers.se).

II. ABSOLUTE CONTINUITY ASSUMPTION

In the statement of [1, Thm. 1], we assumed that $P_{WZ} \ll P_W P_Z$. To avoid measurability issues, we should also assume that $P_W P_Z \ll P_{WZ}$. Similarly, in [1, Thm. 4], we should also assume that $P_{W|\tilde{Z}} P_{\tilde{Z}} P_S \ll P_{W\tilde{Z}S}$.

ACKNOWLEDGEMENTS

The authors would like to thank Pradeep Banerjee, Mahdi Haghifam, Amedeo Roberto Esposito, Max Raginsky, and Yury Polyanskiy who all helped to identify the errors discussed in this note.

REFERENCES

- [1] F. Hellström and G. Durisi, “Generalization bounds via information density and conditional information density,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 824–839, Dec. 2020.
- [2] M. J. Wainwright, *High-Dimensional Statistics: a Non-Asymptotic Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press, 2019.