Random Matrix Methods for Machine Learning

Romain Couillet
University Grenoble Alpes, France
romain.couillet@gipsa-lab.grenoble-inp.fr

Zhenyu Liao
Huazhong University of Science and Technology, China
zhenyu_liao@hust.edu.cn [1]

March 14, 2023

Erratum

Section 3.1.1 "GLRT asymptotics" around Equation (3.2). As a consequence, in order to set a maximum false alarm rate (or false positive, or Type I error) of $r > 0$ in the limit of large n, p, one must choose a threshold $f(\alpha)$ for T_p such that

$$\mathbb{P}(T_p \geq f(\alpha)) = r,$$

that is, such that

$$\mu_{\mathrm{TW}_1}([A_p, +\infty)) = r, \quad A_p = (f(\alpha) - (1+\sqrt{c})^2)(1+\sqrt{c})^{-\frac{4}{3}} c^{\frac{1}{6}} n^{\frac{2}{3}} \tag{3.2}$$

with μ_{TW_1} the Tracy-Widom measure in Theorem 2.15.

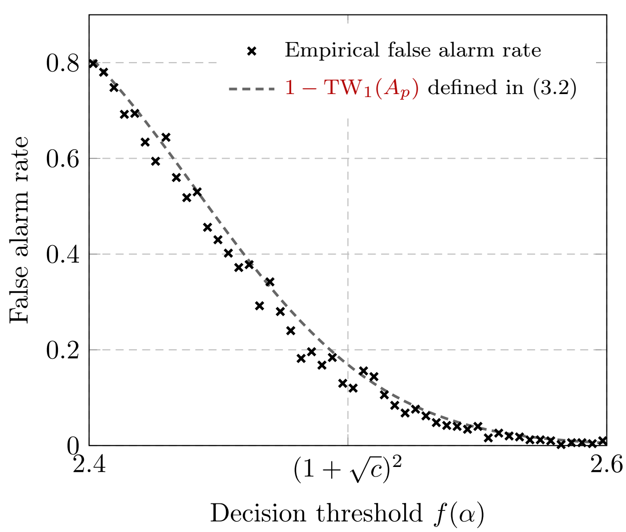

Figure 1: Comparison between empirical false alarm rates and $1 - \mathrm{TW}_1(A_p)$ for A_p of the form in (3.2), as a function of the threshold $f(\alpha) \in [(1+\sqrt{c})^2 - 5n^{-2/3}, (1+\sqrt{c})^2 + 5n^{-2/3}]$, for $p = 256$, $n = 1\,024$ and $\sigma = 1$. Results obtained from 500 runs. Link to code: Matlab and Python.

Section 5.1.1 "Regression with random neural network" after Equation (5.12). The fact that this denominator scales like $\|\gamma\bar{\mathbf{Q}}\|$ as $\gamma \to 0$ explains the major difference between the training and test error behavior in Figure 5.5. Due to the γ^2 prefactor in $\bar{E}_{\text{train}}$, the training error is guaranteed to be finite (even possibly to vanish) as $\gamma \to 0$. But for the test error, since $\gamma\bar{\mathbf{Q}} \to 0$ as N approaches n from each side, if the numerator term $\frac{1}{\hat{n}} \operatorname{tr} \bar{\mathbf{K}}_{\hat{\mathbf{X}}\hat{\mathbf{X}}} - \frac{1}{\hat{n}} \operatorname{tr}(\mathbf{I}_n + \gamma\bar{\mathbf{Q}})(\bar{\mathbf{K}}_{\mathbf{X}\hat{\mathbf{X}}}\bar{\mathbf{K}}_{\mathbf{X}\hat{\mathbf{X}}}^{\mathsf{T}}\bar{\mathbf{Q}})$ does not scale like $\gamma\bar{\mathbf{Q}}$, then $\bar{E}_{\text{test}}$ diverges to infinity at $N = n$. A first counterexample is of course when $\hat{\mathbf{X}} = \mathbf{X}$, for which the numerator term of $\bar{E}_{\text{test}}$ is now

$$\frac{1}{\hat{n}} \operatorname{tr} \bar{\mathbf{K}}_{\hat{\mathbf{X}}\hat{\mathbf{X}}} - \frac{1}{\hat{n}} \operatorname{tr}(\mathbf{I}_n + \gamma\bar{\mathbf{Q}})(\bar{\mathbf{K}}_{\mathbf{X}\hat{\mathbf{X}}}\bar{\mathbf{K}}_{\mathbf{X}\hat{\mathbf{X}}}^{\mathsf{T}}\bar{\mathbf{Q}}) = \frac{\gamma^2}{n} \operatorname{tr} \bar{\mathbf{Q}}\bar{\mathbf{K}}\bar{\mathbf{Q}}$$