Advanced Topics in Machine Learning 2020-2021

Yevgeny Seldin Christian Igel Chloé Rouyer Yi-Shan Wu

Home Assignment 2

Deadline: Tuesday, 15 September 2020, 22:00

The assignments must be answered individually - each student must write and submit his/her own solution. We encourage you to work on the assignments on your own, but we do not prevent you from discussing the questions in small groups. If you do so, you are requested to list the group partners in your individual submission.

Submission format: *Please, upload your answers in a single* `.pdf` *file and additional* `.zip` *file with all the code that you used to solve the assignment. (The* `.pdf` *should* ***not*** *be part of the* `.zip` *file.)*

IMPORTANT: *We are interested in how you solve the problems, not in the final answers. Please, write down the calculations and comment your solutions.*

1 PAC-Bayesian Aggregation (100 points)

In this question you are asked to reproduce an experiment from Thiemann et al. (2017, Section 6, Figure 2) (the paper is an outcome of a master project of the course graduate). Figure 2 corresponds to "the second experiment" in Section 6 "Experimental Results". You are only required to reproduce the experiment for the first dataset, Ionosphere, which you can download from the UCI repository (Asuncion and Newman, 2007). You are allowed to use any programming language you like and any SVM solver you choose. Please, document carefully what you do and clearly annotate your graphs, including legend and axis labels.

Comments:

1. Assuming that you have followed the lectures and read "PAC-Bayesian Analysis" chapter in Yevgeny's lecture notes it should be sufficient to read only the "Experimental Results" section of the paper in order to reproduce the experiment, but you are of course welcome to read the full article.

2. Theorem 6 in the paper corresponds to Theorem 3.32 in Yevgeny's lecture notes, but uses a slightly tighter version of PAC-Bayes-kl inequality. Specifically, the $\ln(n+1)$ term is replaced by $\ln(2\sqrt{n})$, which leads to $\ln((n-r)+1)$ in Theorem 3.32 being replaced by $\ln(2\sqrt{n-r})$ in Theorem 6 in the paper. We do not mind which one you select to work with, but remember to document it in your report.

3. Ideally, you would repeat the experiment several times, say 10, and report the average + some form of deviation, e.g. standard deviation or quantiles, over the repetitions. We have committed a sin by not doing it in the paper. We encourage you to make a proper experiment, but in order to save time you are allowed to repeat the sin (we will not take points for that). Please, do not do that in real papers.

Hint: Direct computation of the update rule for ρ,

$$\rho(h) = \frac{\pi(h)e^{-\lambda(n-r)\hat{L}^{\text{val}}(h,S)}}{\sum_{h'} \pi(h')e^{-\lambda(n-r)\hat{L}^{\text{val}}(h',S)}},$$

is numerically unstable, since for large $n-r$ it leads to division of zero by zero. A way to fix the problem is to normalize by $e^{-\lambda(n-r)\hat{L}^{\text{val}}_{\min}}$, where $\hat{L}^{\text{val}}_{\min} = \min_h \hat{L}^{\text{val}}(h,S)$. This leads to

$$\rho(h) = \frac{\pi(h)e^{-\lambda(n-r)\hat{L}^{\text{val}}(h,S)}}{\sum_{h'}\pi(h')e^{-\lambda(n-r)\hat{L}^{\text{val}}(h',S)}} = \frac{\pi(h)e^{-\lambda(n-r)\left(\hat{L}^{\text{val}(h,S)}-\hat{L}^{\text{val}}_{\min}\right)}}{\sum_{h'}\pi(h')e^{-\lambda(n-r)\left(\hat{L}^{\text{val}}(h',S)-\hat{L}^{\text{val}}_{\min}\right)}}.$$

Calculation of the latter expression for $\rho(h)$ does not lead to numerical instability problems.

Optional Add-on

1. Repeat the experiment with the tandem bound on the majority vote, Theorem 3.38 in Yevgeny's lecture notes, which corresponds to Masegosa et al. (2020, Theorem 9). You can find the details of optimization procedure for the bound in Masegosa et al. (2020, Appendix G). Tandem losses should be evaluated on overlaps of validation sets and n in the bounds should be replaced with the minimal overlap size for all pairs of hypotheses. Compare the results with the first order bound.

Good luck!
Yevgeny, Christian, Chloé, and Yi-Shan

References

Arthur Asuncion and David J. Newman. UCI machine learning repository, 2007. URL `http://archive.ics.uci.edu/ml/index.php`.

Andrés R. Masegosa, Stephan S. Lorenzen, Christian Igel, and Yevgeny Seldin. Second order PAC-Bayesian bounds for the weighted majority vote. Technical report, `https://arxiv.org/abs/2007.13532`, 2020.

Niklas Thiemann, Christian Igel, Olivier Wintenberger, and Yevgeny Seldin. A strongly quasiconvex PAC-Bayesian bound. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, 2017.