
Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2025)

Firstname1 Lastname1^{*1} Firstname2 Lastname2^{*12} Firstname3 Lastname3² Firstname4 Lastname4³
Firstname5 Lastname5¹ Firstname6 Lastname6³¹² Firstname7 Lastname7² Firstname8 Lastname8³
Firstname8 Lastname8¹²

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

(ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.

1. Electronic Submission

References

Author, N. N. Suppressed for anonymity, 2021.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

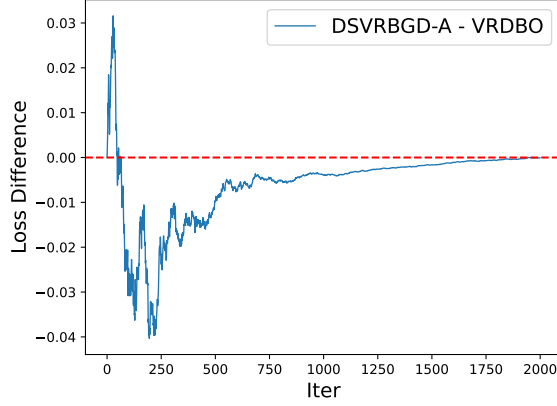
Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

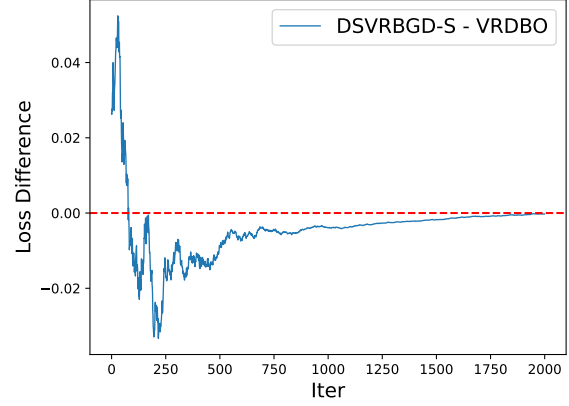
Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R.

^{*}Equal contribution ¹Department of XXX, University of YYY, Location, Country ²Company Name, Location, Country ³School of ZZZ, Institute of WWW, Location, Country. Correspondence to: Firstname1 Lastname1 <first1.last1@xxx.edu>, Firstname2 Lastname2 <first2.last2@www.uk>.

A. New Experimental Results.



(a) $\text{LOSS}_{\text{DSVRBGD-A}} - \text{LOSS}_{\text{VRDBO}}$



(b) $\text{LOSS}_{\text{DSVRBGD-S}} - \text{LOSS}_{\text{VRDBO}}$

Figure 1. The loss difference between our two methods and VRDBO. **A negative difference indicates that our methods converge faster to the stationary point than VRDBO** (The potential reason for the positive difference in the first several steps is that the one-step gradient descent for estimating Hessian-inverse-vector product is not as good as the Neuman series expansion method in the initial stage.). There are 8 workers in this experiment. The training sample's feature x on the k -th worker is generated from a Gaussian distribution $\mathcal{N}(\mu_k, \sigma_k)$, where μ_k is generated from a Uniform distribution $\mathcal{U}(-3, 3)$ and σ_k is generated from a Uniform distribution $\mathcal{U}(1, 25)$.