The Journal of Open Source Software

Candle Optimisers: A Rust crate for optimisation algorithms

Kirpal Grewal [1]

1 Yusuf Hamied Department of Chemistry, University of Cambridge

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 20 December 2023
Published: unpublished

Summary

candle-optimisers is a crate for optimisers written in Rust for use with candle (Mazaré & others (2023)) a lightweight machine learning framework. The crate offers a set of optimisers for training neural networks. This allows network training to be done with far lower overhead than using a full python framework such as PyTorch or Tensorflow.

Statement of need

Rust provides the opportunity for the development of high performance machine learning libraries, with a leaner runtime. However, there is a lack of optimisation algorithms implemented in Rust, with libraries currently implementing only some combination of Adam, AdamW, SGD and RMSProp. This crate aims to provide a set of complete set of optimisation algorithms for use with candle. This will allow Rust to be used for the training of models more easily.

Features

This library implements the following optimisation algorithms:

- SGD (including momentum and Nesterov momentum (Sutskever et al. (2013)))
- RMSprop (Hinton et al. (2012))
- AdaDelta (Zeiler (2012))
- AdaGrad (Duchi et al. (2011))
- AdaMax (Kingma & Ba (2015))
- Adam (Kingma & Ba (2015)) including AMSGrad (Reddi et al. (2018))
- AdamW (Loshchilov & Hutter (2017)) (as decoupled weight decay of Adam)
- NAdam (Dozat (2016))
- RAdam (L. Liu et al. (2019))
- RMSProp (Hinton et al. (2012))
- LBFGS (D. C. Liu & Nocedal (1989))

Furthermore, decoupled weight decay (Loshchilov & Hutter (2017)) is implemented for all of the adaptive methods listed and SGD, allowing for use of the method beyond solely AdamW.

References

Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. *Proceedings of the 4th International Conference on Learning Representations*, 1–4. https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(61), 2121–2159. http://jmlr.org/papers/v12/duchi11a.html

Hinton, G., Srivastava, N., & Swersky, K. (2012). *Neural networks for machine learning*. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Y. Bengio & Y. LeCun (Eds.), *3rd international conference on learning representations, ICLR 2015, san diego, CA, USA, may 7-9, 2015, conference track proceedings*. https://doi.org/10.48550/arXiv.1412.6980

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, *45*(1-3), 503–528. https://doi.org/10.1007/BF01589116

Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2019). On the variance of the adaptive learning rate and beyond. *CoRR*, *abs/1908.03265*. https://doi.org/10.48550/arXiv.1908.03265

Loshchilov, I., & Hutter, F. (2017). Fixing weight decay regularization in adam. *CoRR*, *abs/1711.05101*. https://doi.org/10.48550/arXiv.1711.05101

Mazaré, L., & others. (2023). *Candle: A minimalist ML framework for rust*. https://github.com/huggingface/candle/

Reddi, S. J., Kale, S., & Kumar, S. (2018). On the convergence of adam and beyond. *International Conference on Learning Representations*. https://openreview.net/forum?id=ryQu7f-RZ

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th international conference on machine learning* (Vol. 28, pp. 1139–1147). PMLR. https://proceedings.mlr.press/v28/sutskever13.html

Zeiler, M. D. (2012). ADADELTA: An adaptive learning rate method. *CoRR*, *abs/1212.5701*. https://doi.org/10.48550/arXiv.1212.5701