Importance-aware Data-parallel Training

In importance-aware data parallel training, the dataset is partitioned across the workers based on notions of importance of examples rather than at random. Empirical evaluations have shown that this can lead to better results in terms of model performance compared to “vanilla” data parallelism, at least in image classification tasks.
The aim of this project is to extend the evaluations and experiments to include other tasks such as language modeling, different importance notions, and different partitioning heuristics. We are also interested in studying the potential gains of importance-aware DPT in reducing the aggregated disk I/O and network I/O in distributed training.

Contact: **Sina Sheikholeslami** <sinash@kth.se>
Distributed Computing Group, SCS Division, KTH
Examiner: **Prof. Vladimir Vlassov** <vladv@kth.se>
Distributed Computing Group, SCS Division, KTH

Technical Requirements:

- Familiarity with deep neural networks and distributed training approaches
- Proficiency in implementing techniques and developing libraries using PyTorch

References:

- Sheikholeslami et al. "**The Impact of Importance-Aware Dataset Partitioning on Data-Parallel Training of Deep Neural Networks**." In IFIP International Conference on Distributed Applications and Interoperable Systems, pp. 74-89. 2023.
- Toneva et al. "**An Empirical Study of Example Forgetting during Deep Neural Network Learning**." In *International Conference on Learning Representations*. 2018.
- Katharopoulos and Fleuret. "**Not all samples are created equal: Deep learning with importance sampling.**" In *International conference on machine learning*, pp. 2525-2534. PMLR, 2018.
- Swayamdipta et al. "**Dataset Cartography: Mapping and Diagnosing Datasets with Training Dynamics**." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 9275-9293. 2020.
- Chang et al. "**Active Bias: Training more accurate neural networks by emphasizing high variance samples**." *Advances in Neural Information Processing Systems* 30. 2017.
- Shallue et al. "**Measuring the Effects of Data Parallelism on Neural Network Training**." *Journal of Machine Learning Research* 20, pp. 1-49. 2019.