Joshua Holla (260784204)
Jonathan Maloney-Lebensold (260825605)

Gathering enough data to support Reinforcement Learning in a production setting remains an open challenge. A possible solution could be the construction of a policy derived from the mixing policies from several independent agents with task-relevant experience. The question we would like to answer with this project is whether we can learn a general policy for related environments by combining a Federated Learning approach with the Reinforcement Learning setting.

Federated Learning[1] offers a supervised learning framework that "bring[s] the code to the data, instead of the data to the code"[2]. This distributed machine learning approach leverages data that remains on several nodes (devices or servers) for the task of training a single gradient-based model. After each training round (comprised of a fixed number of epochs) the difference between the current model parameters and the new model parameters is shared and averaged centrally by an aggregator, resulting in a new, better model for the subsequent round of training.
Federated Learning is often used in concert with differential privacy measures. For example, a node may introduce noise in the parameters before sending them to the central aggregator. Extending the Reinforcement Learning with Federated Learning could enable agents to benefit from learners in remote and data-sensitive environments.

We would like to explore the potential benefits of applying Federated Learning to the model-free, gradient-based Reinforcement Learning setting.

Our plan is as follows:

- Achieve state of the art results on DDPG[3] with a set of similar gym environments (for example, altering the physics in a set of Pendulum tasks).
- Throughout training, share network parameters and simulate training rounds.
- Evaluate the impact on the convergence of each agent under this setting.
- Set up another, unseen environment and train a new agent using these general-purpose networks as a starting point.
- Introduce noise as part of the sharing process and evaluate the effect on model convergence.

We hypothesize that by taking an environment and adjusting the setting, that we would still maintain convergence to an acceptable policy and that this averaged policy may provide a good starting point for further generalization.

Joshua Holla (260784204)
Jonathan Maloney-Lebensold (260825605)

# References

1. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Seth, K. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS 17 (pp. 11751191). New York, New York, USA: ACM Press.
https://doi.org/10.1145/3133956.3133982

2. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Roselander, J. (2019). Towards Federated Learning at Scale: System Design. Retrieved from http://arxiv.org/abs/1902.01046

3. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Wierstra, D. (2015). Continuous control with deep reinforcement learning. Retrieved from http://arxiv.org/abs/1509.02971