TD courses:
Optimization for machine learning

09/03/2021 - 18/03/2021

Chuan Xu (chuan.xu@inria.fr)
Othmane Marfoq (othmane.marfoq@inria.fr)

Minitest: 15 mins

1. Explain briefly how backpropagation algorithm is used to compute the gradient of a loss function?

2. Explain briefly what is the synchronous parameter server architecture and its primary disadvantage?

Please send the answers to tareq.si-salem@inria.fr

Our goal:

Train ML model using multiple computing units (GPUs or CPUs) from the cluster

Our schedule:

1st course: Usage of NEF, PyTorch, package for distributed communication
2nd course: Build synchronous Parameter-Server architecture, evaluation part
3rd course: Federated learning - implement FedAvg algorithm
4th course: Advanced techniques in federated learning

Evaluation: participation and the small tasks to finish after classes.

Usage of NEF: https://wiki.inria.fr/ClustersSophia/Clusters_Home

- Hardware details [https://wiki.inria.fr/ClustersSophia/Hardware]
- Check available nodes and job state
  [https://nef-frontal.inria.fr/monika]
- FAQ [https://wiki.inria.fr/ClustersSophia/FAQ_new_config]

log in account: ssh user@nef-frontal.inria.fr

Reserve CPUs:

```
oarsub -l /nodes=1/core=2,walltime=3 -I
```

(reserving 2 cores from one node for three hours)

Reserve GPUs:

```
oarsub -p "gpu='YES' and gpucapability>='5.0'" -l /gpunum=1,walltime=1 -I
```

Preparing PyTorch environment:

```
module load conda/5.0.1-python3.6
conda create --name virt_pytorch
conda activate virt_pytorch
module load cuda/9.2
module load cudnn/7.1-cuda-9.2
module load gcc/7.3.0
module load mpi/openmpi-2.0.0-gcc
module load pytorch/1.4.0
```

1st TD: Usage of NEF, PyTorch, package for distributed communication