

USER GUIDE

1. Executing Serial Python Scripts

All Python scripts can simply be executed using the command:

```
python SerialBCART.py
```

2. Executing Spark Scripts in Local Mode

The parallel BCART scripts must be saved in the root directory of the Spark folder (/usr/local/spark). Similarly, the script must be executed from the Spark folder, using:

```
./bin/spark-submit ParallelBCART.py 512 1000 20 1.5 1 results_folder
```

The last 6 elements of this command are the command line arguments for the parameters of the script, being the:

- Number of trees (int),
- Number of iterations (int),
- Minimum Split Value (int),
- Beta (float),
- Number of threads (int) and
- File name to save the results (String).

3. Executing Spark Scripts in Cluster Mode

Once the code has been copied over to the Hartree computers, the following command will execute the code:

```
spark-submit --master = yarn --deploy-mode cluster --num-executors=12  
--executor-cores=10 ParallelBCART.py 512 1000 20 1.5 1 results_folder
```

To run in local mode, omit “--deploy-mode cluster”.