Data Science Toolbox Portfolio Questions 10 Parallel Algorithms

Daniel Lawson — University of Bristol

Block 10

Portfolio 10

Choose **one question** and write up to **one page** about it. You are free to conduct further experiments to add weight to your results, and any additional material you generate can be submitted as an appendix. See The Assessment Page for advice.

These questions may make reference to the content from the current block.

Question R10.1: By extending the benchmarking from Block 9 Workshop (09.3) to include parallel code as provided in Workshop 10.3, provide examples of *parallel speedup* in which a) the *efficiency* is 1, and b) the efficiency is lower than 1 but still of value (i.e. the parallel algorithm does more overall compute than the sequential but is quicker). These should be algorithms for which the *computational efficiency* exhibits these features - they may have constant terms that make practice harder. Focus your writeup on the choice and scaling of the algorithms.

Question R10.2: Investigate Spark (e.g. using pyspark or sparkR) and implement a simple mapping-and-reducing problem, providing the code as an appendix and writing up in the format of a tutorial.

Question R10.3: Explain the difference between Matrix Multiplication as implemented on a CPU vs a massively parallel GPU, from the paper Understanding the Efficiency of GPU Algorithms for Matrix-Matrix Multiplication. In terms of concepts we've covered in DST, what is the take-home message?