Data Science Toolbox Question Sheet

10.1 Parallel algorithms

Daniel Lawson

Block 10

1. Explain at a high level what a computational graph is, and how it can be used in parallel processing.
2. Given the compute time of a parallel algorithm as $T_p = (1 - P)T_s + PT_s/S$, what is P, S and T_S? Define parallel speedup. What is the maximum speedup?
3. Give two examples of problems that can be solved using an "embarrassingly parallel" approach.
4. Describe vectorisation of code and explain why it is a good idea.
5. Explain what an accumulation/reduction computation is. How can this be parallelised? What is the fastest that it can be achieved in parallel?
6. Give a high level description of the map/reduce computational framework.