

Big Data Analytics

Exercise Sheet 4

Prof. Dr.Dr. Lars Schmidt-Thieme, Mohsan Jameel

Information Systems and Machine Learning Lab University of Hildesheim

May 12th, 2016

Submission until May 18th, 2016, to mohsan.jameel@ismll.de

Exercise 1: MPI Collective Communication (7 points)

- a) Why collective communication algorithms are required? What are the benefits of using built-in collective routines over implementing from scratch?
- b) A parallel program running with 8 processes. The **process 1** needs to send an array of **size 40** integers to other processes such that after completion of the communication each process has exactly **5** elements of this array. How can you achieve such communication using collective communication routine? Write down the signature of the method/routine or if it is not possible with existing methods please provide the logic.
- c) Let say a parallel programing is running with 4 processes. The **process 0** needs to send and array of **size 40** to other processes. It is however to be noted that after completion of communication **process 0** will have **5** elements, **process 1** will have **8** elements, **process 2** will have **12** elements and **process 3** will have **15** elements. How can you achieve such communication using collective communication routine? Write down the signature of the method/routine or if it is not possible with existing methods please provide the logic.

Hint: Please check the MPI APIs and documentation.

Exercise 2: Logistic Regression with MPI (8 points)

- a) Referring to the lecture slides (tutorial lecture 04), a generic sketch of parallel algorithm is provided. Your task is to provide parallel algorithm that is based on MPI framework i.e. you have to specify which MPI function call is used for communication and at which part of the algorithm. For example either you will use **Bcast** or **Scatter** or **Alltoall** to communication training and test data.
- b) Reference to serial code provided, your task is to implement parallel Logistic Regression using MPI framework.