Convex Optimization for Big Data
Implementing Shotgun in the "Cloud"

Kunal Ghosh, 546247 // Jussi Ojala, 544605

September 13, 2016

1   Objective

The objective of this assignment is to implement the "Shotgun" algorithm for parallel coordinate descent. We already had a Matlab implementation as a part of the course work and In this part of the course we aim to do the following.

- Implement the "Shotgun" algorithm in Apache Spark using the pySpark API.

- Document the setup of Amazon Elastic Map Reduce (EMR) cloud infrastructure with all the necessary libraries to run pySpark code.

- Document the parts of the algorithm which were not available in pyspak.mllib.linalg.distributed module and we had to implement ourselves. This is to ensure these code snippets can be used in other projects.

- Run the pySpark Implementation of the "Shotgun" algorithm on a few large datasets and do preliminary analysis of the results.

To summarize, we want this assignment to be a guide on how to **speed-up optimization algorithm on the cloud** rather how to implement one particular optimization algorithm "Shotgun" using Apache spark.