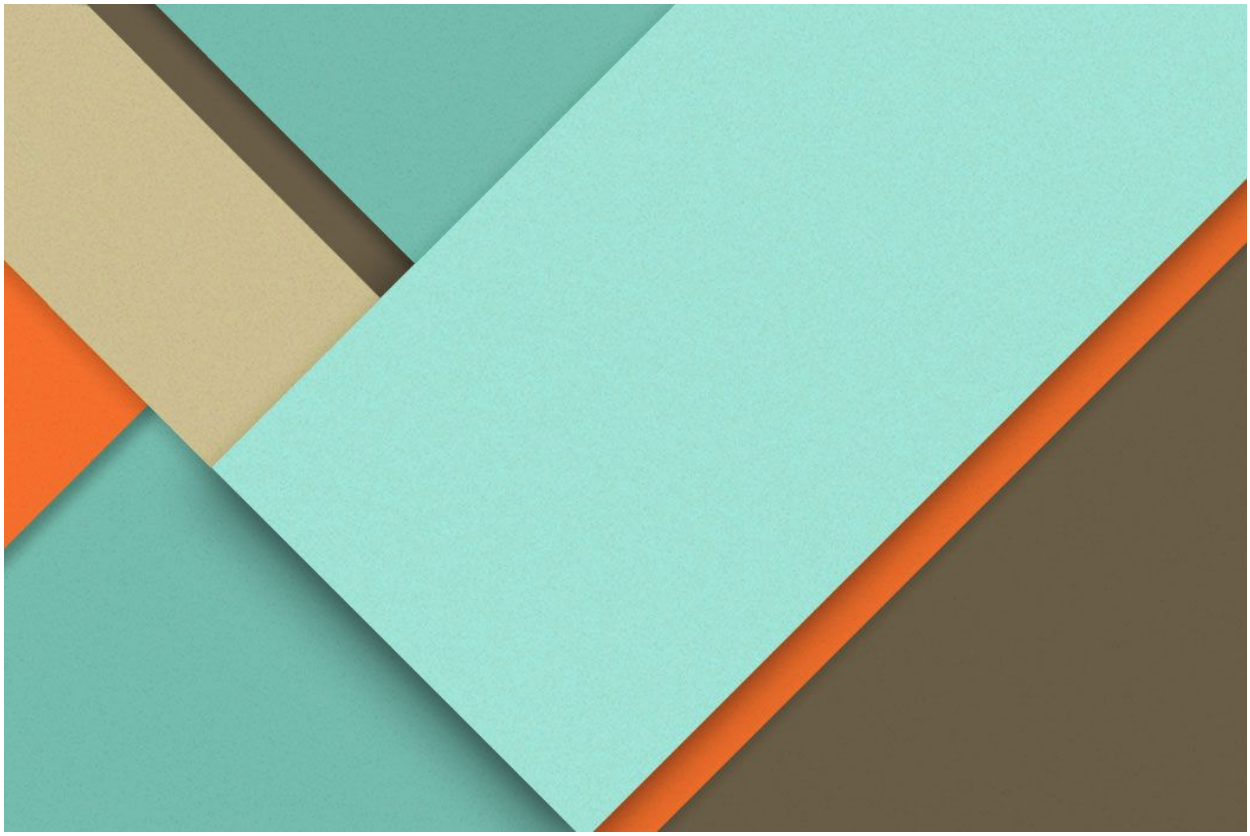

COMS4995: Project Proposal

MAP-REDUCE Library

11.22.2019

Kanishk Vashisht, Sambhav Anand
Kv2295, sa3433

Overview

We are proposing a generic Map-Reduce Library that abstracts away the implementation from the application programmer. Our project will focus on implementing it on a single machine where our program will make use of available cores on a machine to parallelize the job.

Key Points of Focus

1. The library should be very high level for any application programmer and the actual implementation should be abstracted away.
2. The programmer should be able to simply specify the mappers and reducers as simple functions. The monad's that we use to implement the library should do all the job of spreading the work across different cores and the reduction portion.
3. We should be able to build an application on top of this library for demo purposes. Some ideas include distributing the job of finding support vectors across different jobs, simple but very fast wordCount, fast implementation of tf-idf etc.
4. The focus is going to be on saving time as compared to a sequential implementation and being able to scale to large inputs easily.

Project Goals

1. A developer defines mappers and reducers and calls the library on them. The library distributes the sequential inputs and then maps over them and handles the final reduction.
2. We want to implement a distributed SVM using our library as our demo