

# Distributed Machine Learning Mini Batch K-Means

Rohit Nair, Abhijit Karanjkar

December 1, 2016

We have referred the research paper on Text Categorization [1] and Web-scale k-means clustering [2]. [3] was a good resource for TF, IDF and cosine similarity calculations. HARP guides were found in [4].

## References

- [1] Y.; Rose T.; Lewis, D. D.; Yang and F Li. A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5:361–397, 2004.
- [2] D. Sculley. Web-scale k-means clustering, in: Proceedings of the 19th international conference on world wide web. page 1177–1178, 2010.
- [3] Jana Vembunarayanan. *Tf-Idf and Cosine similarity*. 2013.
- [4] K.Y. Yip; D.W. Cheung; M.K. Ng. Harp: A practical projected clustering algorithm. pages 1041–4347, 2004.