
K-Means for Big Data

Reuben K. McCreanor
Department of Statistics
Duke University
Durham, NC 27708
`reuben.mccreanor@duke.edu`

Wei (Emily) Shao
Department of Statistics
Duke University
Durham, NC 27708
`wei.shao@duke.edu`

Abstract

ADD ABSTRACT HERE

1 Introduction

2 Discussion of Methods

3 Results

3.1 Predicting Airline Delay

References

- [1] American Statistical Association. *Data Expo 2009*. Statistical Computing and Statistical Graphics. Accessed October 20, 2015. <http://stat-computing.org/dataexpo/2009/the-data.html>.
- [2] National Oceanic and Atmospheric Administration. *Air Traffic*. Science on a Sphere. Accessed October 30, 2015. <http://sos.noaa.gov/Datasets/dataset.php?id=44>.
- [3] Ball, Michael et. al. *Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States*. The National Center of Excellence for Aviation Operations Research (NEXTOR), October 2010.
- [4] Socrata. *Airport Codes mapped to Latitude and Longitude in the United States*. Open Data. Accessed November 10, 2015. <https://opendata.socrata.com/dataset/Airport-Codes-mapped-to-Latitude-Longitude-in-the-rxrh-4cxm>.
- [5] United States Department of Transportation. *On-Time Performance*. Bureau of Transportation Statistics. Accessed November 15, 2015. http://www.transtats.bts.gov/Fields.asp?Table_ID=236.
- [6] American Statistical Association. *Data expo 2009 Supplementary Data*. Statistical Computing and Statistical Graphics. Accessed November 14, 2015. <http://stat-computing.org/dataexpo/2009/supplemental-data.html>.
- [7] Link for Shiny App: https://reuben.shinyapps.io/Airline_Delay_Comparison.
- [8] United States Department of Transportation. *Understanding the Reporting of Causes of Flight Delays and Cancellations*. Bureau of Transportation Statistics. Accessed November 20, 2015. <http://www.rita.dot.gov/bts/help/aviation/html/understanding.html>.
- [9] Federal Aviation Administration. *Types of Delay*. Accessed November 18, 2015 http://aspmhelp.faa.gov/index.php/Types_of_Delay.

4 Appendix

4.1 OLS, LASSO and Ridge Regression Methods