Assignment: 04 - Linear Regression

Bucket: 1 - Homework

Due Date: 2016-02-12

Grading Hidden? no

Teams? yes

Description:

The price of a ticket depends, in part, on the amount of fuel consumed in the particular itinerary (a factor of distance and winds). Furthermore, prices have a tendency to increase over time. To understand airline pricing, computer a simple linear regression that models the cost of tickets for different airlines. Give a new ranking of airlines with respect to price.

- Group assignment, two students
- Create a pipeline of jobs to run on EMR (map-reduce) and locally (R, maybe Bash scripts).
    - Determine the airlines active in 2015.
    - Work with the flights by those airlines in 2010-2014.
    - For each of those airlines compute linear regressions for {distance traveled, flight time} to price from 2010-2014.
    - Automatically generate graphs (with R) showing a linear fit of the variable to price for each airline.
- Include your graphs in your report, and conclude which airline is cheapest.
- Is distance traveled or flight time a better variable? Why?

To submit:

- Source code and build scripts.
- A report (PDF).
- A readme with build and execute instructions (text).

Don't submit:

- jars, class files, word documents, etc.
- Amazon keys.

Assignment Download: ()

Your Submissions

Team Members: Yogiraj Awati, Sarita Joshi, Ashish Kalbhor, Sharmodeep Sarkar

New Team Submission (/assignments/42/submissions/new)

Date	Status	Automatic	Teacher	Score	Link
2016-02-12 23:55:07 -0500		ø / 100	100.0 / 100	100.0 / 100	View (/submissions/3455)

Course Page (http://www.ccs.neu.edu/home/ntuck/courses/2016/01/cs6240/index.html) | Piazza (https://piazza.com/class/ij4yepvnz8v3zf)

ajax-status: none