

Final Project Proposal

Kevin Chen
jiafengchen@college.harvard.edu

Yufeng Ling
yufengling@college.harvard.edu

Francisco Rivera
frivera@college.harvard.edu

November 15, 2017

1 Problem Description

The primary objective of this project is to predict trip duration from taxi data published by the TLC in New York City. An accurate predictor has a host of implications, among them,

- Accurate predictions can allow NYC residents to give themselves a smaller time buffer when commuting in the city, and thus save time.
- Estimates of trip duration play a central role in the pricing of trips by ride-sharing apps, which increasingly feel pressured to quote a price before the completion of the trip.
- From the perspective of a city planner, easing congestion requires an understanding of which trips systematically take longer.

2 The Data

Talk about the data that we get here. Columns we observe, time period we get, and how much data there is.

3 Methodology

3.1 Latent Variables Model

