Predicting Cab Fare

Initial Proposal

How much will the ride cost?

It's always nice to know how much one's cab fare will cost. Especially when wondering about comparisons with Uber and Lyft. There are also concerns with how much one wants to spend, and expects to spend.

There is a Kaggle Competition currently underway that asks exactly this question. The dataset consists of 55 million rows of New York cab rides. Included columns are the timestamp of the pick-up, in addition to the latitude and longitude of the pick-up and drop-off. The presents a very realistic scenario: you know where you are, where you're going, and what time it is.

Cab fare predictions can be utilized in any geographical area. My clients are regular people trying to decide whether they should take a cab.

I will use deep learning, in particular, keras regression, to solve this problem. I will start by exploring a variety of nodes and layers, in addition to epochs and batch-sizes. As for the data itself, I will create new columns that will help determine traffic and costs such as rush hour, the location of busy areas like Manhattan, the taxicab distance between places, and the time of day.

I will present my code, a slide deck, a blog, and my results in the Kaggle Competition. The competition ends in approximately one month, on September 25.

Corey Wade
August 29, 2018