Predicting Cab Fare
Initial Proposal

How much will the ride cost?

Determining the cost of cab fares in advance can only be advantageous. In an era where Uber, Lyft, and public transportation costs are often known in advance, the same is needed for cab fares.

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.

My clients are regular people trying to decide whether they should take a cab. Machine learning can help solve the problem. The final pipeline may be added to any website or mobile application utilizing transportation industries.

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, 2018.

Corey Wade
August 29, 2018