Taxi Challenge Update

What’s done:

1. Data modification in Jupiter
2. First model and training result
3. Result is not quite normal distribution and loss is still decreasing

TODO：

1. Detailed report and all decisions
2. Examine why Train and Validation has identical loss value
3. Edit Jupiter Notebook to contain text descriptions for each step
4. Examine test data to make a prediction
5. Improve the model base on the result
6. Create local python files

Current Problem:

1. Fast convergence with High loss for Train vs Cross Validation data

- Training with a much smaller Learning rate with more iteration. Running test set tonight

Potential Improvement:

1. More outliers (Passenger Counts etc.)
2. Better way to eliminate outliers
3. I chosed Duration, Speed, Fare Amount, Trip Distance, Rush Hour/ OverNight, add or remove feature and test result
4. More Hidden layers?

Question for challenge:

1. Master and slave and how to use them effectively
2. How to do a better Feature Engineering and figure out what works for the best? Maybe a Pair plot (Possible to Reverse engineer for selecting Features? looking at data figure and see if there is a recognizable pattern)
3. Unsupervised Learning problems?
4. Commonly used Libraries? (TensorFlow)Book recommendations and what next course
5. Commonly used VM for spark (GCP AWS, local or other)
6. Next Possible Challenge <https://www.microsoft.com/en-us/download/confirmation.aspx?id=54765> Dogs vs Cats recognition
7. Best way to find learning rate other than try error
8. Popular optimizer to use? (Adam, Gradient Decent, etc.)