

Datasets

The training and testing datasets were provided by the Kaggle competition and can be found on the competition website (<https://www.kaggle.com/c/new-york-city-taxi-fare-prediction>)

Notebooks

- 1) NY_taxi_feature_Cleaning_Engineering_Scaling.ipynb
- 2) Deep_Neural_Network_Regression.ipynb

Two .ipynb notebooks as mentioned above are also provided. The order of execution of these notebooks is 1) and then 2). Hence, one can execute the 1) notebook after downloading the competition dataset in the current directory and it takes a total of ~ 3-5 mins to execute 1) notebook. Afterwards for 2) notebook, instead of training the DNN model again, one can load the best weights for this DNN model from the file name 'best_weights.hdf5', which is also provided, thus saving the training time. Additionally, I am also providing the 'submission.csv' file that was uploaded to the Kaggle competition to get the final score.

Relevant software and libraries used

NY taxi feature Cleaning Engineering Scaling.ipynb:

```
import numpy as np
import pandas as pd
import os
from sklearn.preprocessing import StandardScaler
from sklearn.externals import joblib
```

Deep Neural Network Regression.ipynb:

```
%matplotlib inline
from matplotlib.pyplot import figure, show
from sklearn.model_selection import train_test_split
import pandas as pd
import os
import numpy as np
from sklearn import metrics
from scipy.stats import zscore
import tensorflow as tf
from keras.models import Sequential
from keras.layers.core import Dense, Activation, Dropout
from keras.callbacks import EarlyStopping
from keras.callbacks import ModelCheckpoint
import matplotlib.pyplot as plt
from keras.layers.normalization import BatchNormalization
```

```
from sklearn.preprocessing import StandardScaler
from sklearn.externals import joblib
from keras.optimizers import Adam
```