WWT: (Producing Fig 24)

* It tracks # of daily views of Wikipedia articles from July 1st, 2015 to December 31st, 2016.
* There are 4 dataset splits real dataset (training and test) + generated dataset (training and test); generated test dataset are not used at all.
* Each of these 4 splits has two arrays: (features) and (attributes/classes); Attributes array are not used at all in this regression task.
* Each timestamp has only one measurement: how many views/
* Each of these feature arrays is split into X and Y. X has a shape of (50000, 500, 1) and y has a shape of (50000, 50, 1).
* Some web pages have > 2000 page views, others have <10.
* It is just a time-series forecasting task.
* The regression task: Given the first 500 consecutive days page count views of this Wikipedia article, predict the page count views for the next 50 days.
* Timestamp: The date the page views are counted on.
* Metrics used to evaluate: R2 score (coefficient of determination)
* Jupyter Notebook Dir: '/rdata/yelnady/DoppelGANger'
* Generated Data Config: aux\_disc-True,dataset-web,epoch-400,epoch\_checkpoint\_freq-1,extra\_checkpoint\_freq-5,run-0,sample\_len-10,self\_norm-True,vis\_num\_sample-10,
* Tensorflow: MLP-1layer: One Dense Layer (100,activation='relu'),Adam(learning\_rate=0.001),loss='mean\_squared\_error' ;

MLP-5Layes: Five Dense Layers (200,activation='relu'),Adam(learning\_rate=0.001),loss='mean\_squared\_error'

* sklearn: Linear Regression: LinearRegression(), Kernel Ridge: KernelRidge()
* Results:

