Evaluation Results

Overview

Datasets of varying sequence length are created and sequential rnn models trained using them.

Sequence lengths of 150, 300, 400, 600 are used.

Train and test data are created independently.

Either ‘train’ or ‘test’ in the dataset name e.g ‘sequence_length_150__num_examples_10000_test_1_dp’ specifies which data it is.

Validation data is set to test data in the training pipeline. In this way the rmse loss and mse metric validation curves corresponds to the model performance on test data.

Metrics curve also includes baseline of 0.1767 to compare.

Moreover, after each training 20 randomly sampled examples are selected from test data and their ground truth, predicted value and rmse is printed.

A separate script called evaluate.py can also be used independently after training is done to evaluate model on number of examples specified as a parameter.

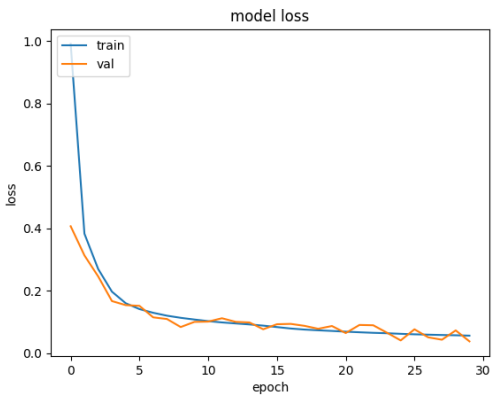
Results:

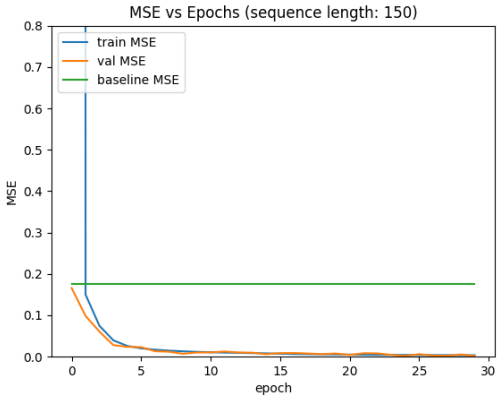
Sequence length: 150

Dataset train = datasets/sequence_length_150__num_examples_100000_train_1_dp.json

Dataset test = datasets/sequence_length_150__num_examples_10000_test_1_dp.json

Trained model = trained_models/rnn__sequence_length_150__examples_100k_epochs_30





Evaluated on 20 randomly chosen examples from test set.

```
label: 0.90 , prediction = 0.91, mse = 0.00013
label: 1.40 , prediction = 1.43, mse = 0.00083
label: 1.00 , prediction = 1.01, mse = 0.00015
label: 1.20 , prediction = 1.21, mse = 0.00019
label: 0.70 , prediction = 0.72, mse = 0.00046
label: 1.10 , prediction = 1.11, mse = 0.00022
label: 0.60 , prediction = 0.61, mse = 0.00015
label: 0.20 , prediction = 0.20, mse = 0.00000
label: 1.10 , prediction = 1.11, mse = 0.00012
label: 1.10 , prediction = 1.11, mse = 0.00018
label: 0.80 , prediction = 0.83, mse = 0.00087
label: 0.70 , prediction = 0.73, mse = 0.00074
label: 0.70 , prediction = 0.72, mse = 0.00060
label: 1.10 , prediction = 1.12, mse = 0.00050
label: 1.30 , prediction = 1.31, mse = 0.00022
label: 0.50 , prediction = 0.51, mse = 0.00019
label: 0.80 , prediction = 0.81, mse = 0.00010
label: 0.30 , prediction = 0.33, mse = 0.00073
label: 1.20 , prediction = 1.22, mse = 0.00060
label: 0.50 , prediction = 0.51, mse = 0.00006
```

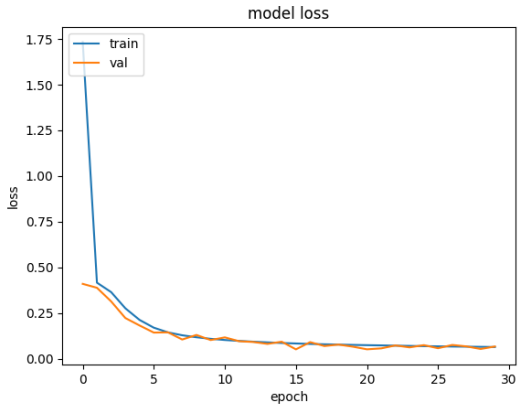
It can be seen from the results that model is not overfitting.

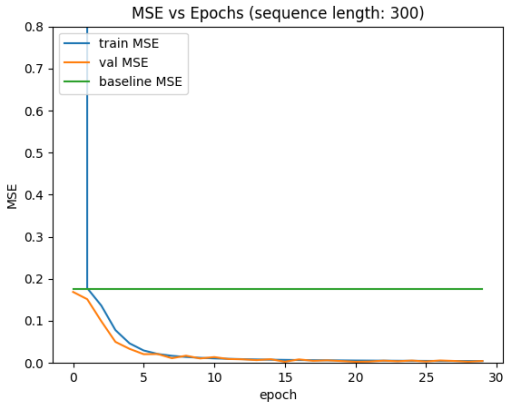
Sequence length: 300

Dataset train = datasets/sequence_length_300__num_examples_100000_train_1_dp.json

Dataset test = datasets/sequence_length_300__num_examples_10000_test_1_dp.json

Trained model = trained_models/rnn__sequence_length_300__examples_100k_epochs_30





Evaluated on 20 randomly chosen examples from test set.

```
label: 1.30 , prediction = 1.40, mse = 0.00927
label: 1.00 , prediction = 1.08, mse = 0.00589
label: 1.00 , prediction = 1.06, mse = 0.00410
label: 1.00 , prediction = 1.06, mse = 0.00377
label: 1.20 , prediction = 1.23, mse = 0.00086
label: 0.70 , prediction = 0.76, mse = 0.00349
label: 0.90 , prediction = 0.95, mse = 0.00271
label: 1.20 , prediction = 1.24, mse = 0.00146
label: 0.70 , prediction = 0.78, mse = 0.00582
label: 1.10 , prediction = 1.21, mse = 0.01143
label: 0.60 , prediction = 0.68, mse = 0.00574
label: 0.20 , prediction = 0.24, mse = 0.00190
label: 0.50 , prediction = 0.55, mse = 0.00241
label: 0.60 , prediction = 0.66, mse = 0.00367
label: 0.70 , prediction = 0.78, mse = 0.00593
label: 1.20 , prediction = 1.29, mse = 0.00816
label: 1.00 , prediction = 1.05, mse = 0.00263
label: 1.10 , prediction = 1.15, mse = 0.00289
label: 0.80 , prediction = 0.87, mse = 0.00495
label: 1.10 , prediction = 1.15, mse = 0.00232
```

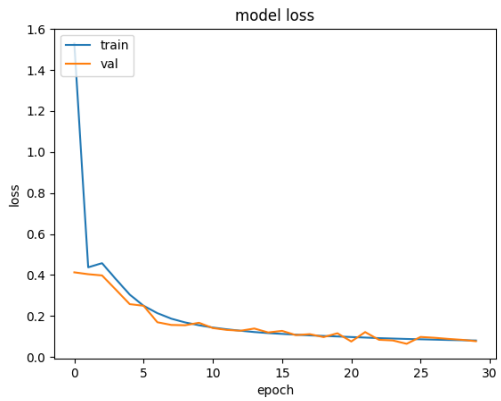
It can be seen from the results that model is not overfitting.

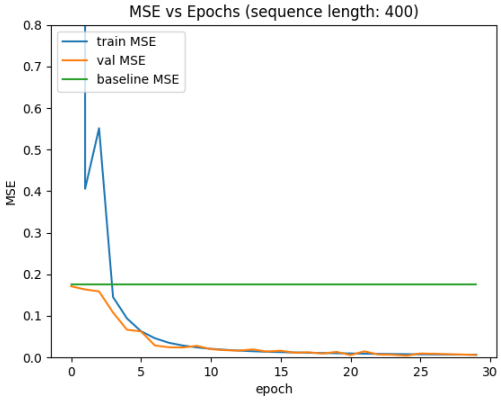
Sequence length: 400

Dataset train = datasets/sequence_length_400__num_examples_100000_train_1_dp.json

Dataset test = datasets/sequence_length_400__num_examples_10000_test_1_dp.json

Trained model = trained_models/rnn__sequence_length_400__examples_100k_epochs_30





Evaluated on 20 randomly chosen examples from test set.

```
label: 0.80 , prediction = 0.85, mse = 0.00215
label: 1.20 , prediction = 1.24, mse = 0.00192
label: 1.70 , prediction = 1.74, mse = 0.00194
label: 1.50 , prediction = 1.50, mse = 0.00001
label: 1.00 , prediction = 1.07, mse = 0.00530
label: 1.20 , prediction = 1.29, mse = 0.00892
label: 0.20 , prediction = 0.28, mse = 0.00719
label: 0.40 , prediction = 0.46, mse = 0.00366
label: 1.00 , prediction = 1.01, mse = 0.00017
label: 1.30 , prediction = 1.36, mse = 0.00420
label: 1.10 , prediction = 1.07, mse = 0.00078
label: 0.50 , prediction = 0.57, mse = 0.00501
label: 1.20 , prediction = 1.24, mse = 0.00198
label: 1.40 , prediction = 1.50, mse = 0.01044
label: 1.70 , prediction = 1.71, mse = 0.00018
label: 1.10 , prediction = 1.13, mse = 0.00090
label: 1.30 , prediction = 1.35, mse = 0.00264
label: 1.30 , prediction = 1.39, mse = 0.00752
label: 1.30 , prediction = 1.37, mse = 0.00454
label: 0.60 , prediction = 0.76, mse = 0.02696
```

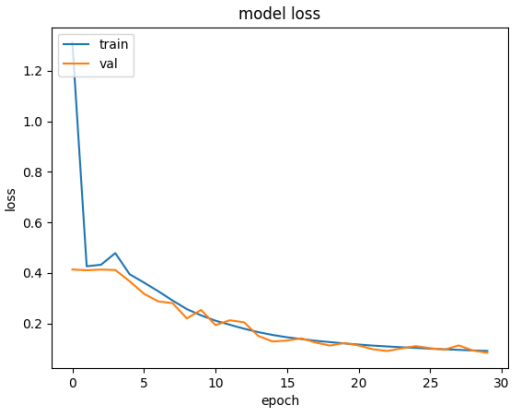
It can be seen from the results that model is not overfitting.

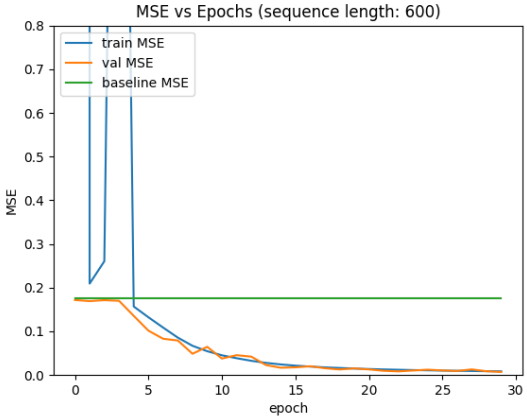
Sequence length: 600

Dataset train = datasets/sequence_length_600__num_examples_100000_train_1_dp.json

Dataset test = datasets/sequence_length_600__num_examples_10000_test_1_dp.json

Trained model = trained_models/rnn__sequence_length_600__examples_100k_epochs_30





Evaluated on 20 randomly chosen examples from test set.

```
label: 0.70 , prediction = 0.62, mse = 0.00602
label: 1.80 , prediction = 1.72, mse = 0.00595
label: 0.40 , prediction = 0.33, mse = 0.00511
label: 0.60 , prediction = 0.50, mse = 0.00996
label: 1.30 , prediction = 1.31, mse = 0.00011
label: 1.10 , prediction = 1.04, mse = 0.00381
label: 1.10 , prediction = 1.08, mse = 0.00031
label: 1.00 , prediction = 0.98, mse = 0.00053
label: 1.70 , prediction = 1.65, mse = 0.00252
label: 1.00 , prediction = 0.82, mse = 0.03364
label: 1.40 , prediction = 1.41, mse = 0.00005
label: 1.00 , prediction = 0.82, mse = 0.03421
label: 1.40 , prediction = 1.33, mse = 0.00548
label: 0.80 , prediction = 0.73, mse = 0.00482
label: 1.00 , prediction = 0.88, mse = 0.01388
label: 0.20 , prediction = 0.18, mse = 0.00024
label: 0.70 , prediction = 0.62, mse = 0.00721
label: 0.50 , prediction = 0.43, mse = 0.00552
label: 0.80 , prediction = 0.68, mse = 0.01323
label: 0.60 , prediction = 0.48, mse = 0.01533
```

It can be seen from the results that model is not overfitting.