Fixed Time-Step Numerical Sequence Learning

...

Online Encyclopedia of Integer Sequences (OEIS)

Github repository link: https://github.com/HeTalksInMaths/DeepMLProj-OEIS

Project Overview

Recurrent Neural Networks (RNNs) can learn from arbitrary length inputs to model subsequent values.

Consider prediction at specific time-step.

Predict 10th sequence value in subset of OEIS. Multiple approaches.

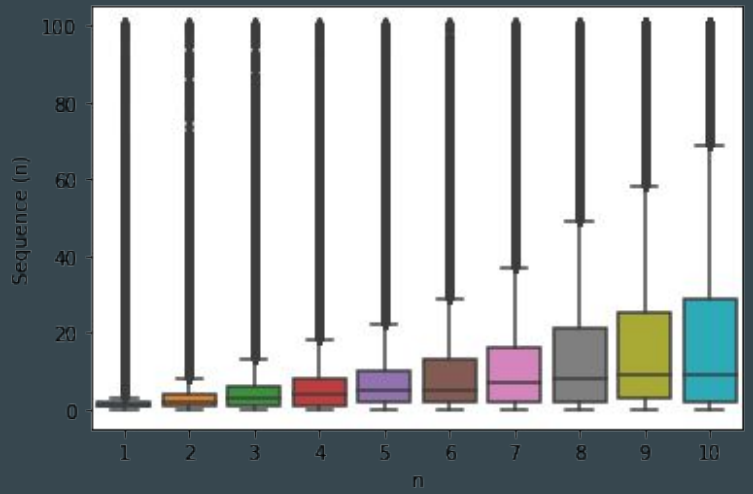
Applications - stock market price on specified date, booking curve projection etc.

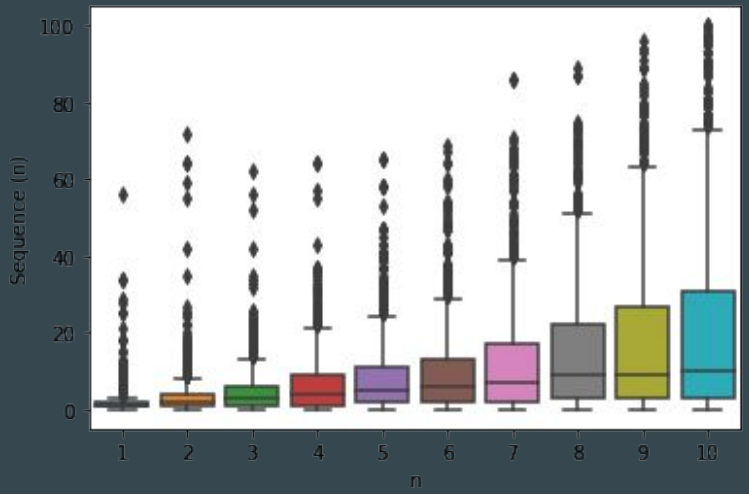
Data

Online Encyclopedia of Integer Sequences (OEIS), 350,000 + seqs.

1000 sequences sampled with first 10 values between 0 and 100.

Not drawn from underlying distribution, but some apparent structure.





Experimental Setup

Method 1:

Train - 700 seqs., val. - 200 seqs., test - 100 seqs.

From 1st - 9th seq. values, learn 10th seq. value.

Method 2:

Use all 1000 seqs. Train - from 1st - 7th seq. values, learn 8th seq. value.

Val - from 1st - 8th seq values, learn 9th seq. value.

Test - same as before.

Model Architecture

Simple RNNs as sequence are short.

Keras implementation - fully connected.

Hyperparameter tuning: 10, 20, 50, 100, 200, 500 units.

Mean absolute error loss, early-stopping to prevent overfitting.

Results

Method 1

Units	Mean Absolute Error (Validation)
10	5.396
20	5.199
50	4.901
100	4.661
200	4.921
500	4.556

Method 2

Units	Mean Absolute Error (Validation)
10	6.161
20	5.119
50	5.332
100	4.635
200	4.706
500	4.531

Method	Mean Absolute Error (Test)
Benchmark	5.270
1	3.440
2	5.444

All figures 4s.f.

Key Takeaways (for OEIS)

Fitting to 10th sequence value leads to better generalizability across sequences. Even with only 1000 sampled sequences.

Next step -consider other / larger samples of sequences or different time-steps. Expand to different data sets.