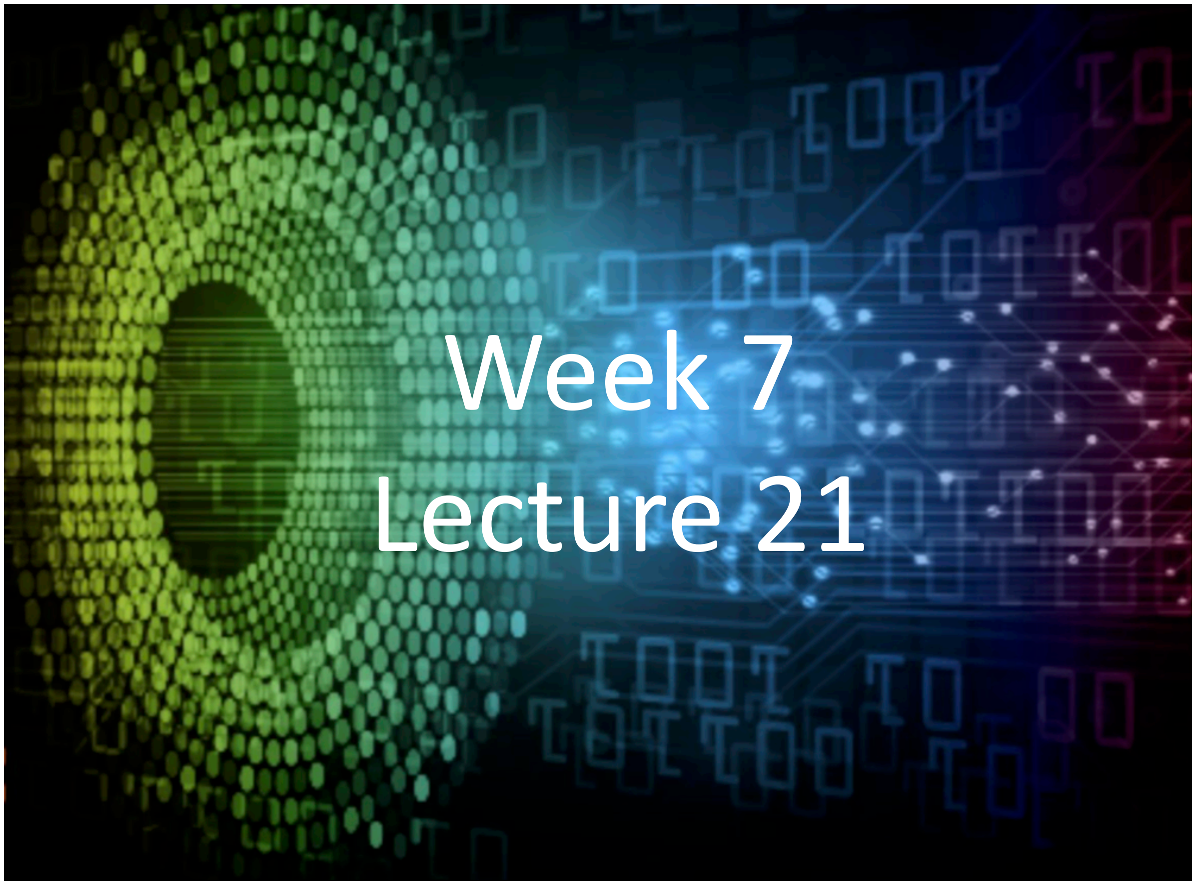
Introduction to Deep Learning Applications and Theory



ECE 596 / AMATH 563

Previous Lecture: Advanced Applications of RNNs

- Multivariate Timeseries Applications
  - Preparation
  - Filtering
  - Completion

This Lecture: Advanced Applications of RNNs

- Multivariate Timeseries Applications
  - Prediction
  - Forcasting
  - Translation
  - Reconstruction
  - Anomaly Detection

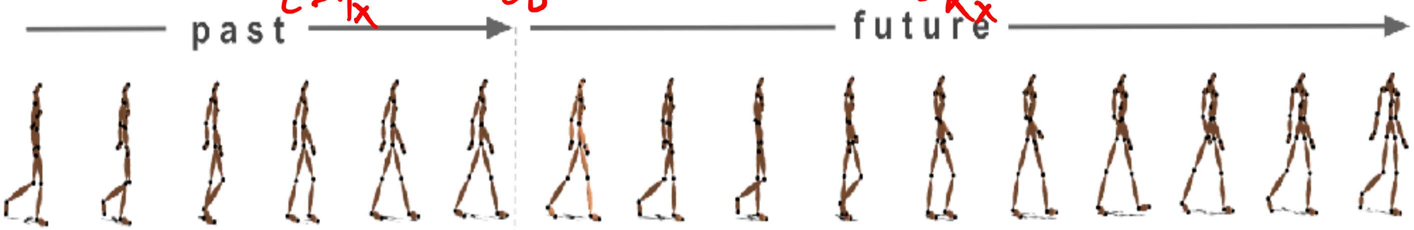
Prediction Setup

Prediction

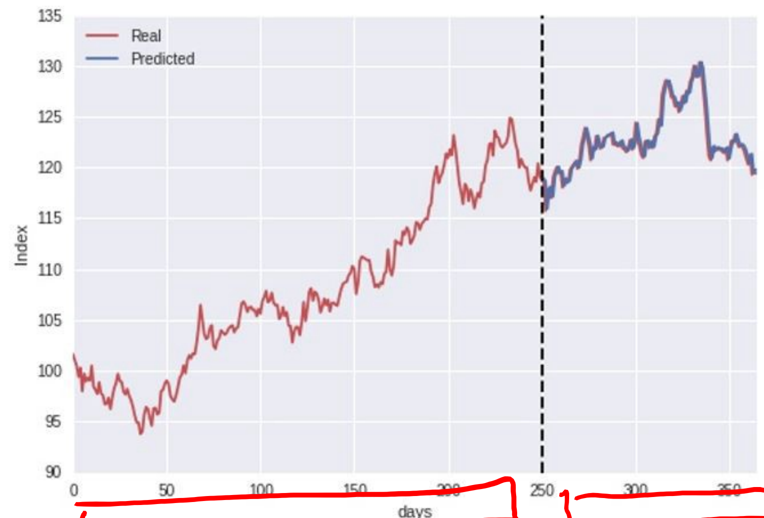
joints $\boxed{\text{past}}$ $\boxed{\text{Future}}$

$t-T_x$ t_0 $t+K_x$

complex system



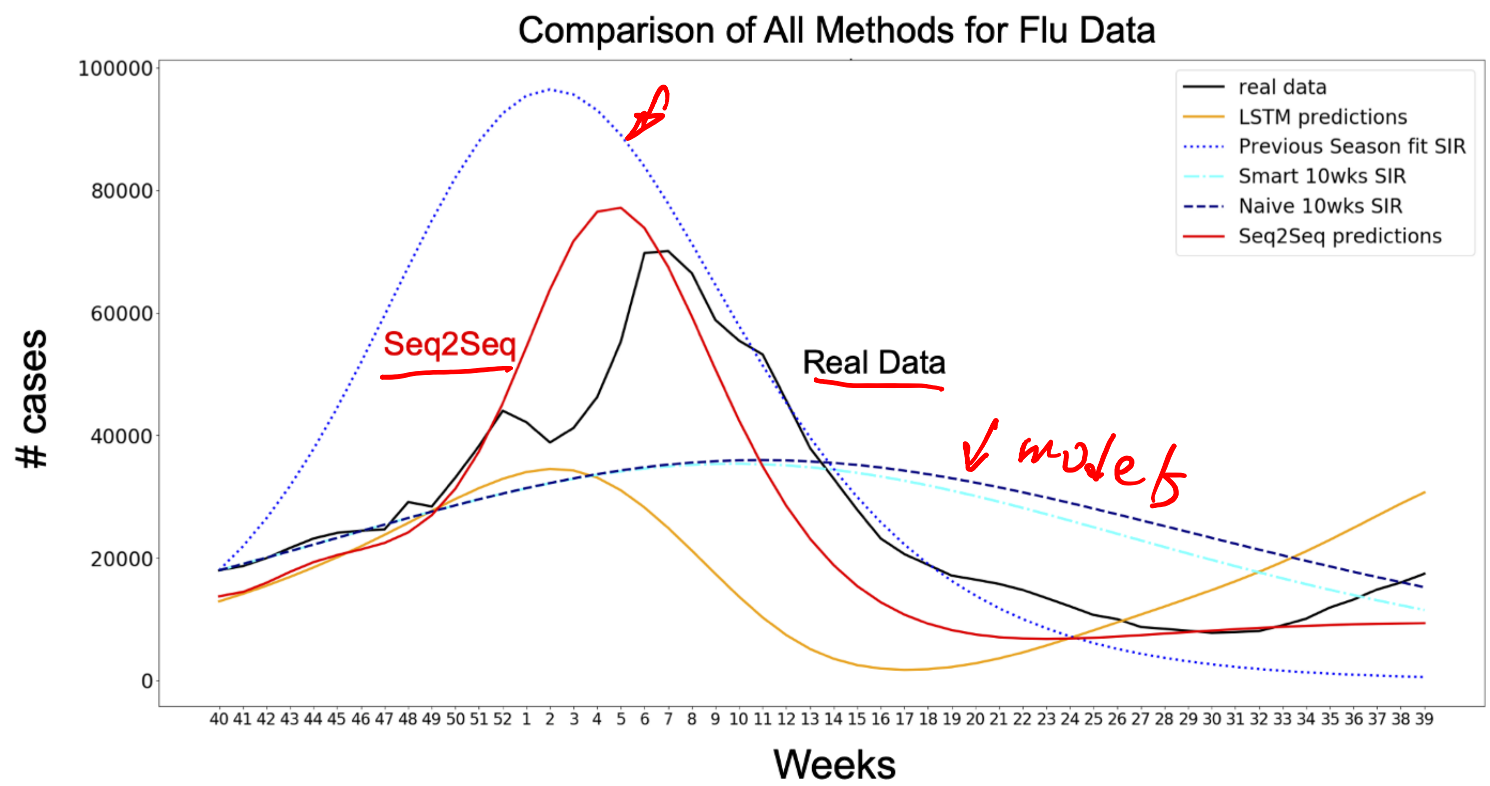
Stochastic system



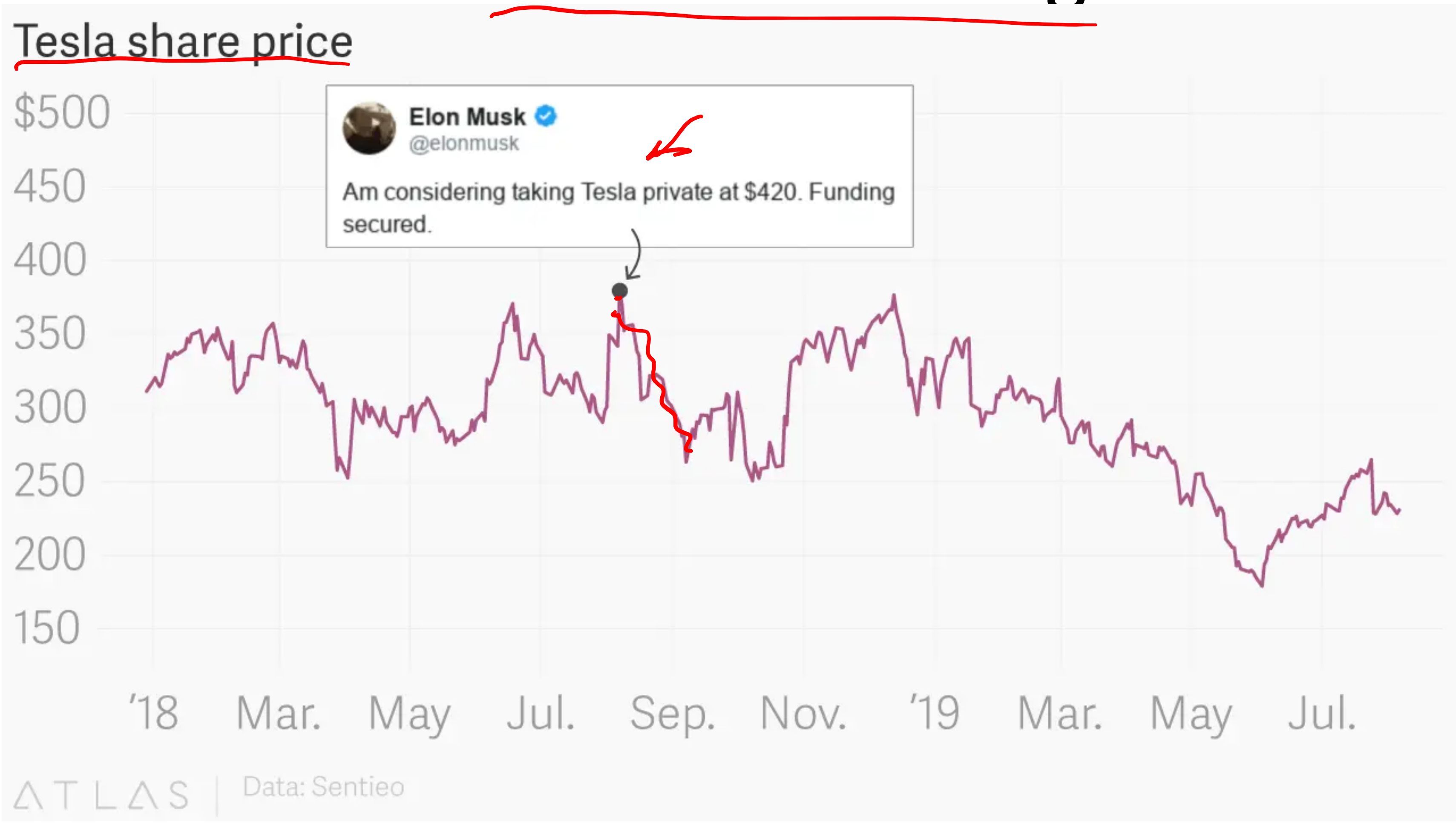
Past

Future

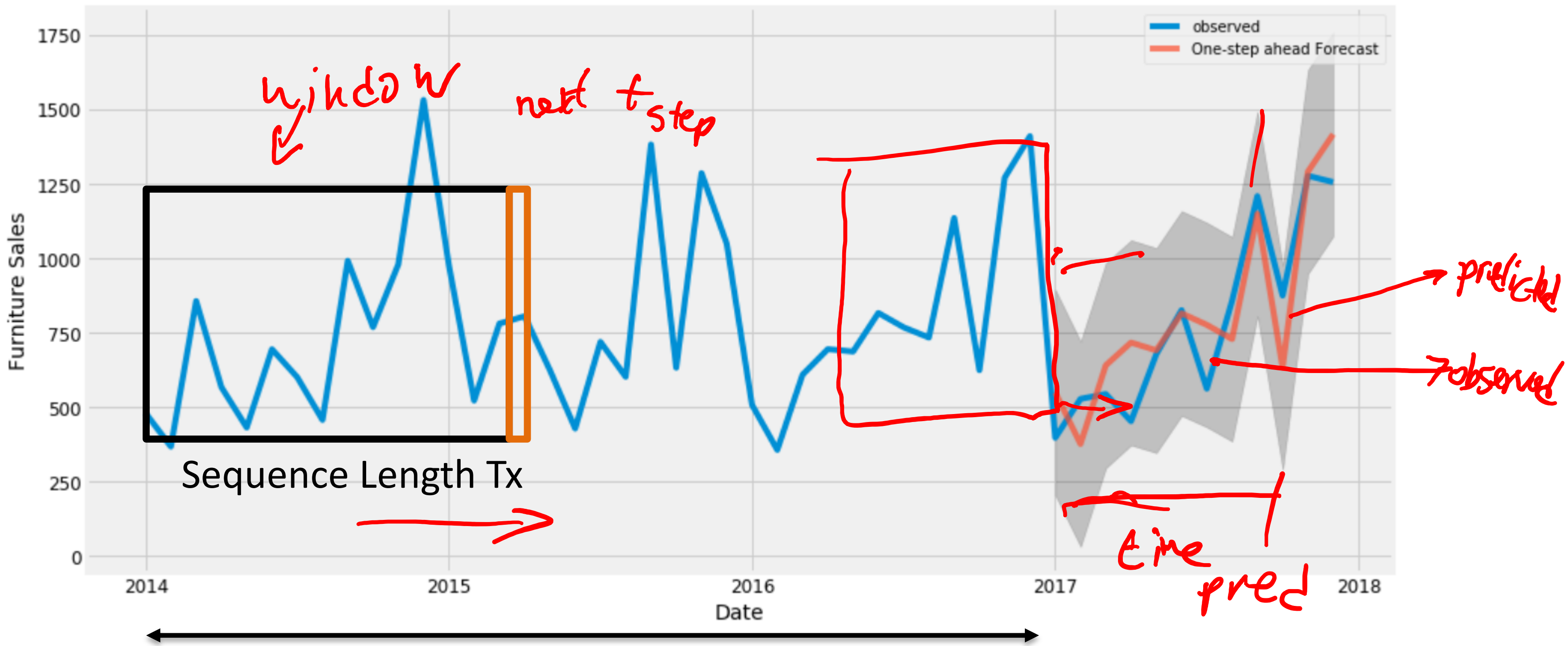
Model vs. Data driven



Prediction -> Forcasting -> Decision Making



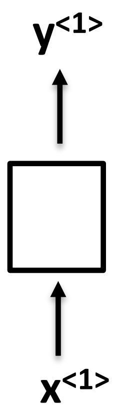
Moving Prediction - Forcasting



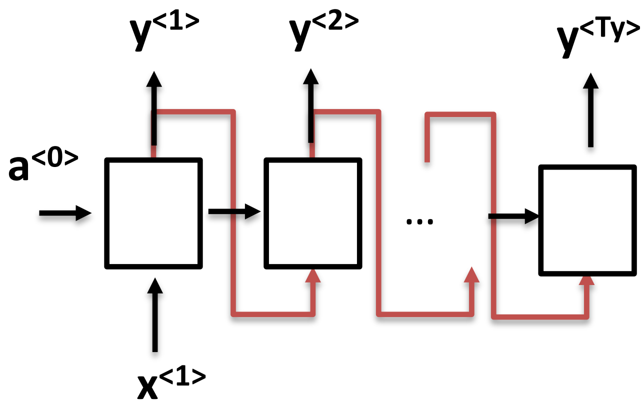
Training

Testing

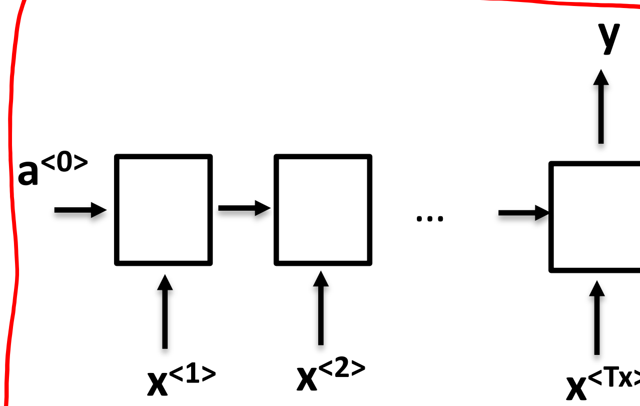
Types of RNNs



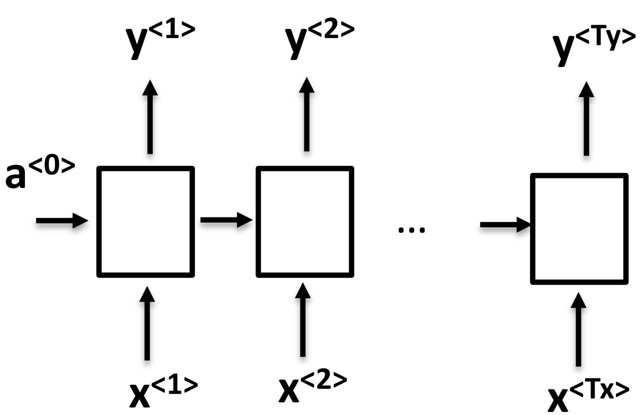
One to one



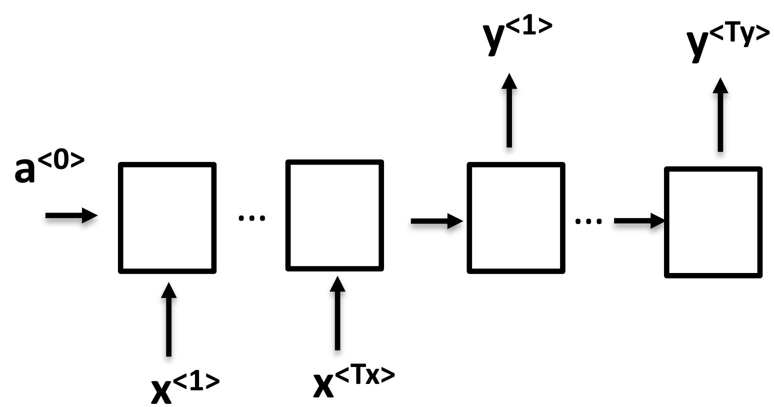
One to many



Many to one

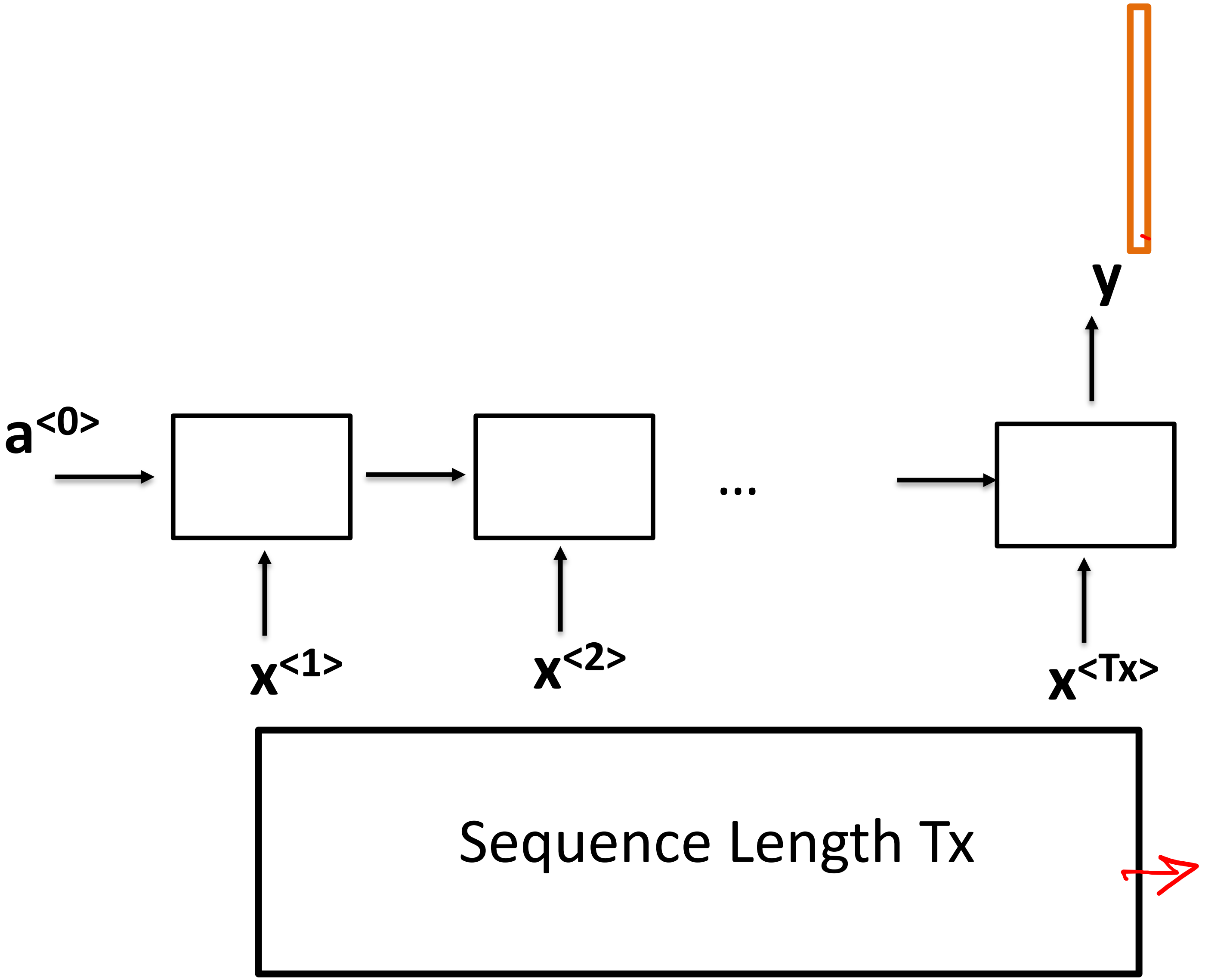


Many to many

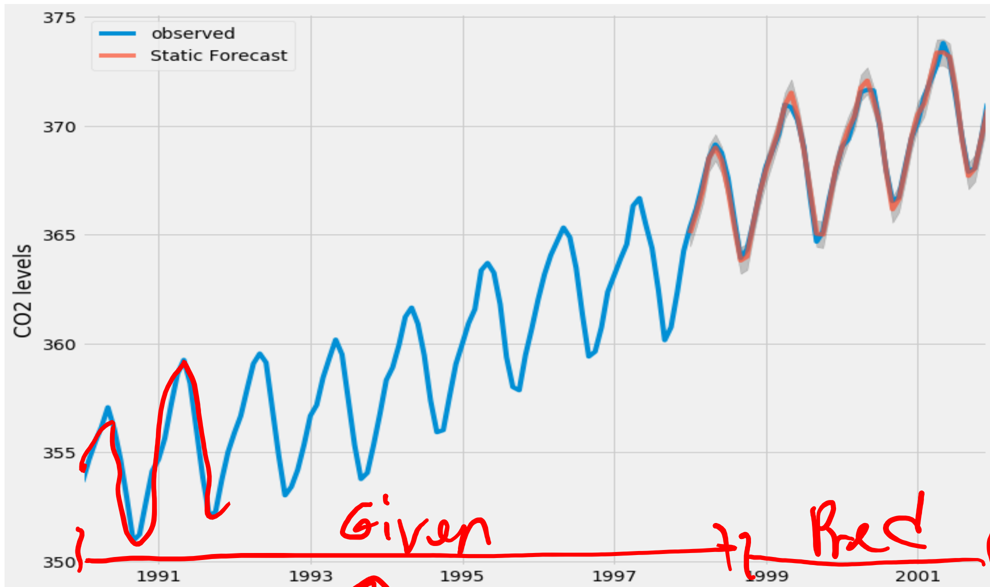


Many to many

One Step Prediction



Long-Term Prediction (Translation)



Input

Output

LTP

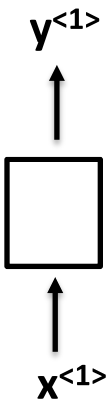
MT

«Hey Siri, où puis-je acheter une Tesla?»	"Hey Siri, where can I buy a Tesla?"

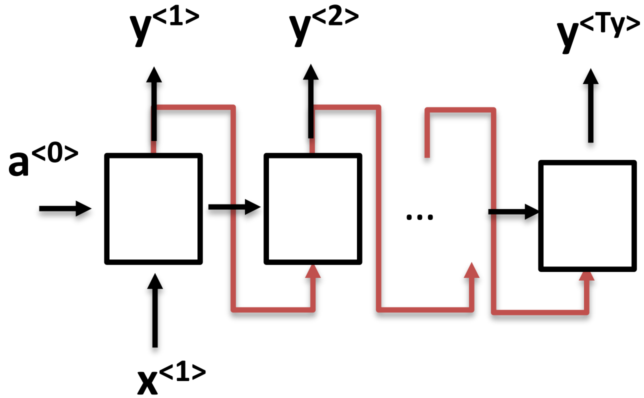
input

output

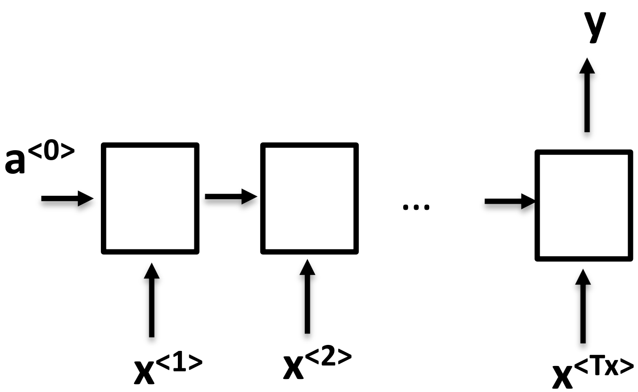
Types of RNNs



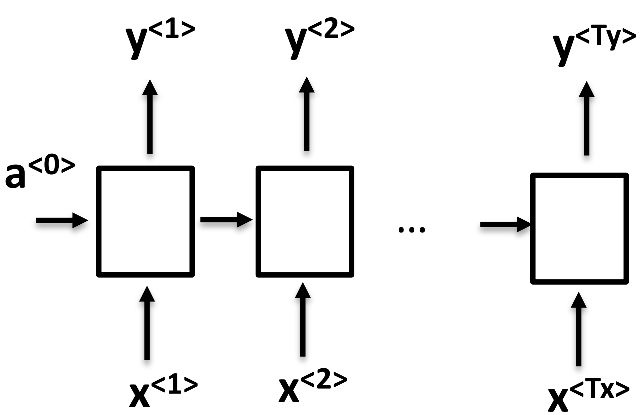
One to one



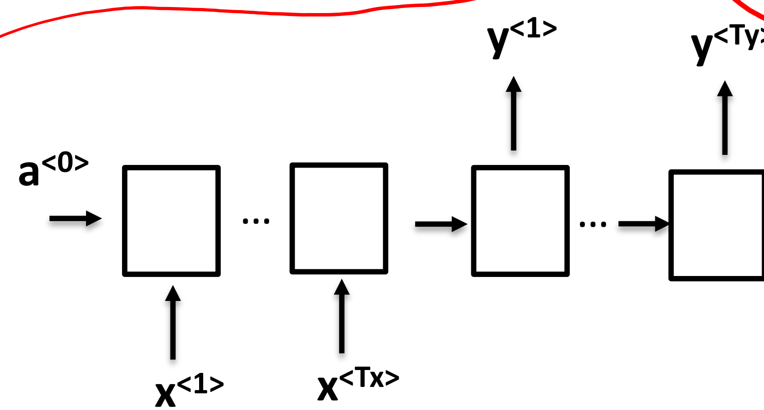
One to many



Many to one

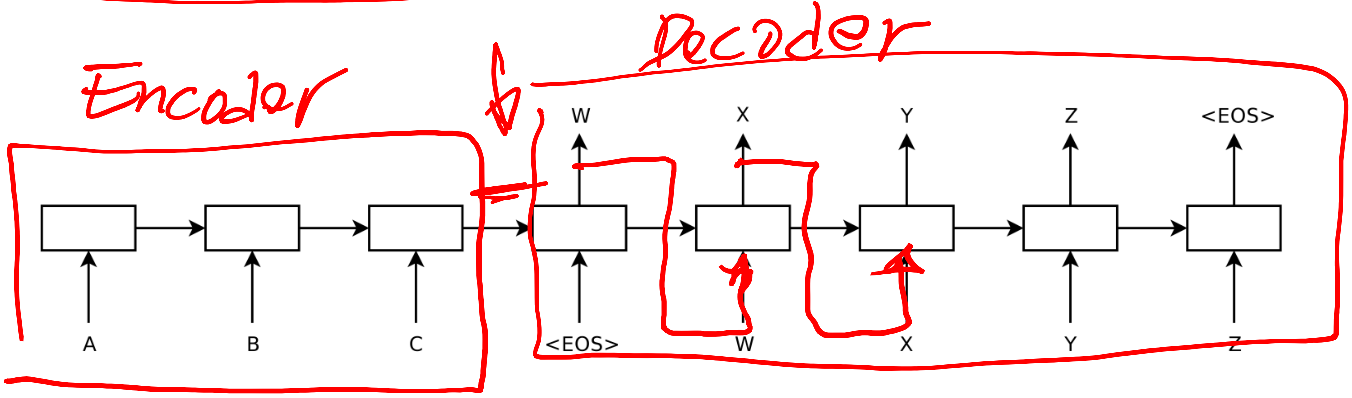


Many to many



Many to many

Sequence to Sequence (Seq2Seq)



ABC -> WXYZ

Computer Science > Computation and Language

[Submitted on 10 Sep 2014 (v1), last revised 14 Dec 2014 (this version, v3)]

Sequence to Sequence Learning with Neural Networks

Ilya Sutskever, Oriol Vinyals, Quoc V. Le

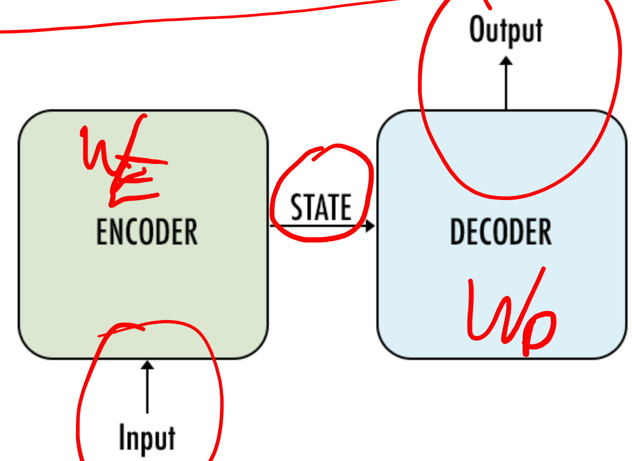
Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector. Our main result is that on an English to French translation task from the WMT'14 dataset, the translations produced by the LSTM achieve a BLEU score of 34.8 on the entire test set, where the LSTM's BLEU score was penalized on out-of-vocabulary words. Additionally, the LSTM did not have difficulty on long sentences. For comparison, a phrase-based SMT system achieves a BLEU score of 33.3 on the same dataset. When we used the LSTM to rerank the 1000 hypotheses produced by the aforementioned SMT system, its BLEU score increases to 36.5, which is close to the previous best result on this task. The LSTM also learned sensible phrase and sentence representations that are sensitive to word order and are relatively invariant to the active and the passive voice. Finally, we found that reversing the order of the words in all source sentences (but not target sentences) improved the LSTM's performance markedly, because doing so introduced many short term dependencies between the source and the target sentence which made the optimization problem easier.

Sequence to Sequence

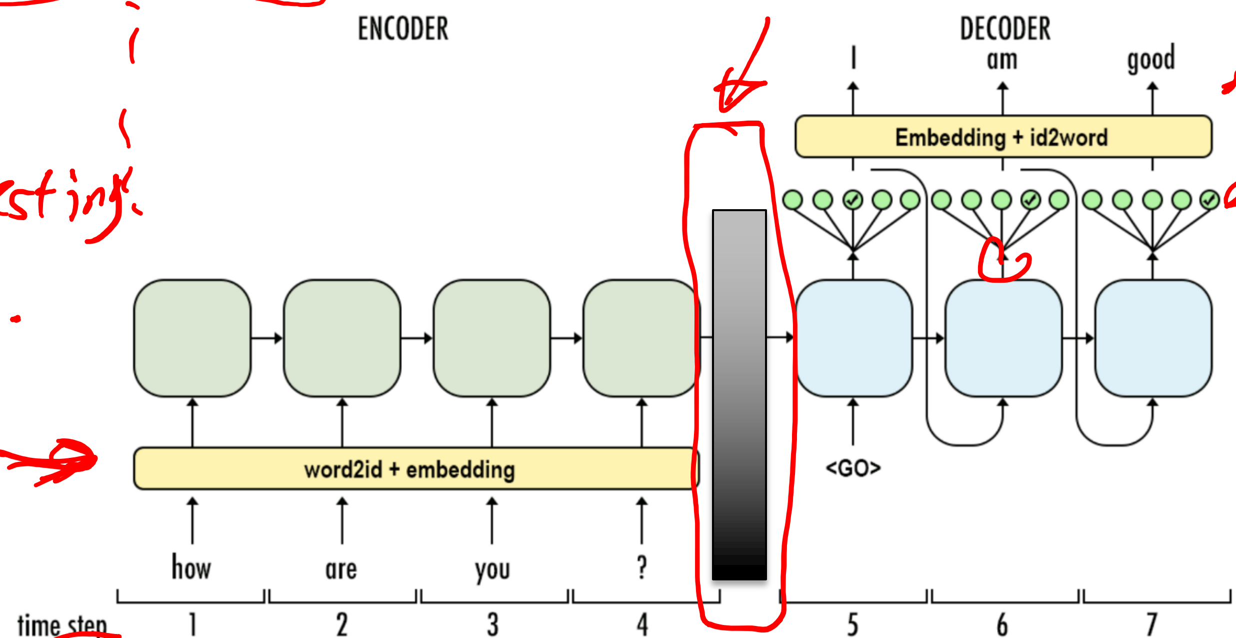
Training

input

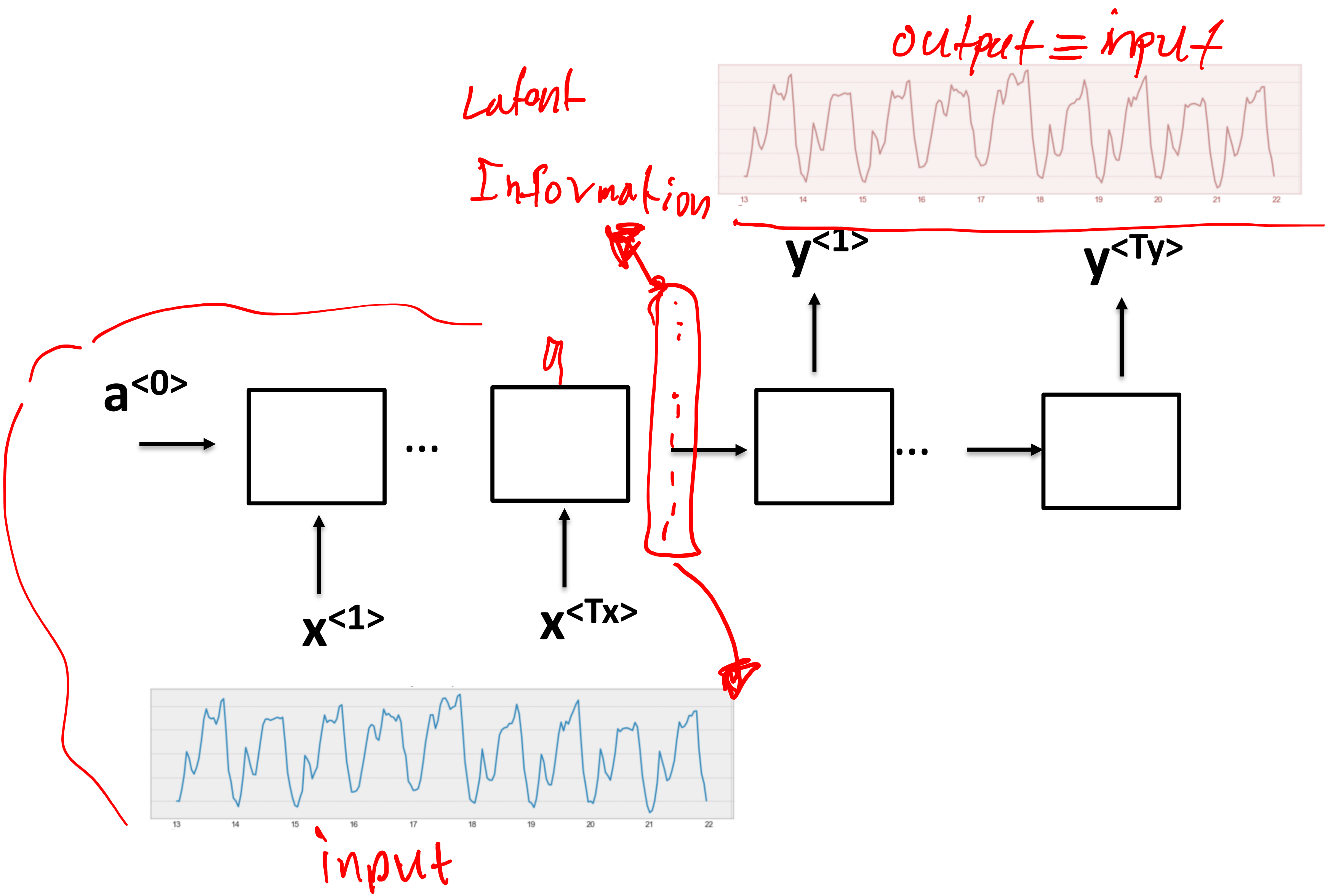
Output



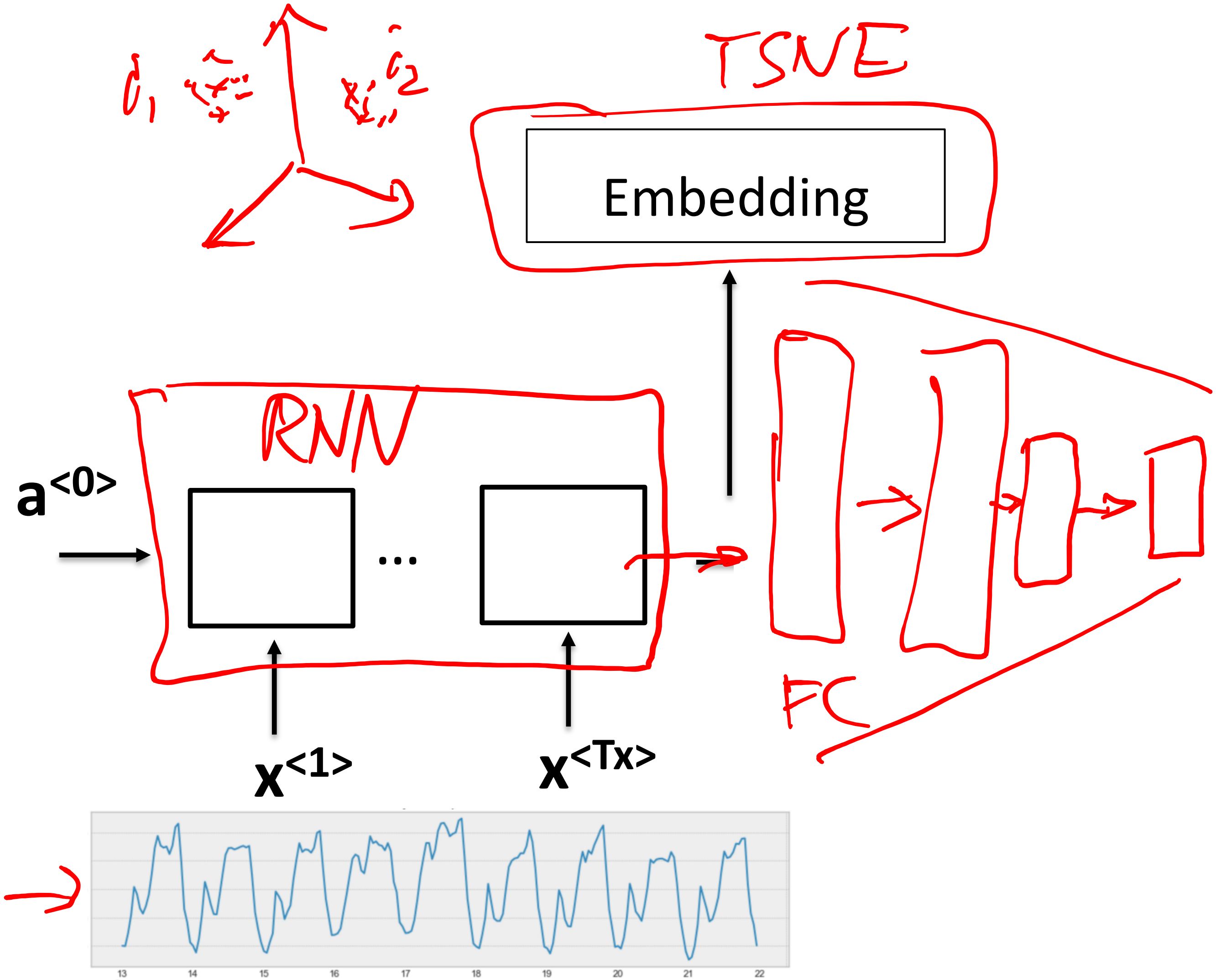
Testing



Reconstruction

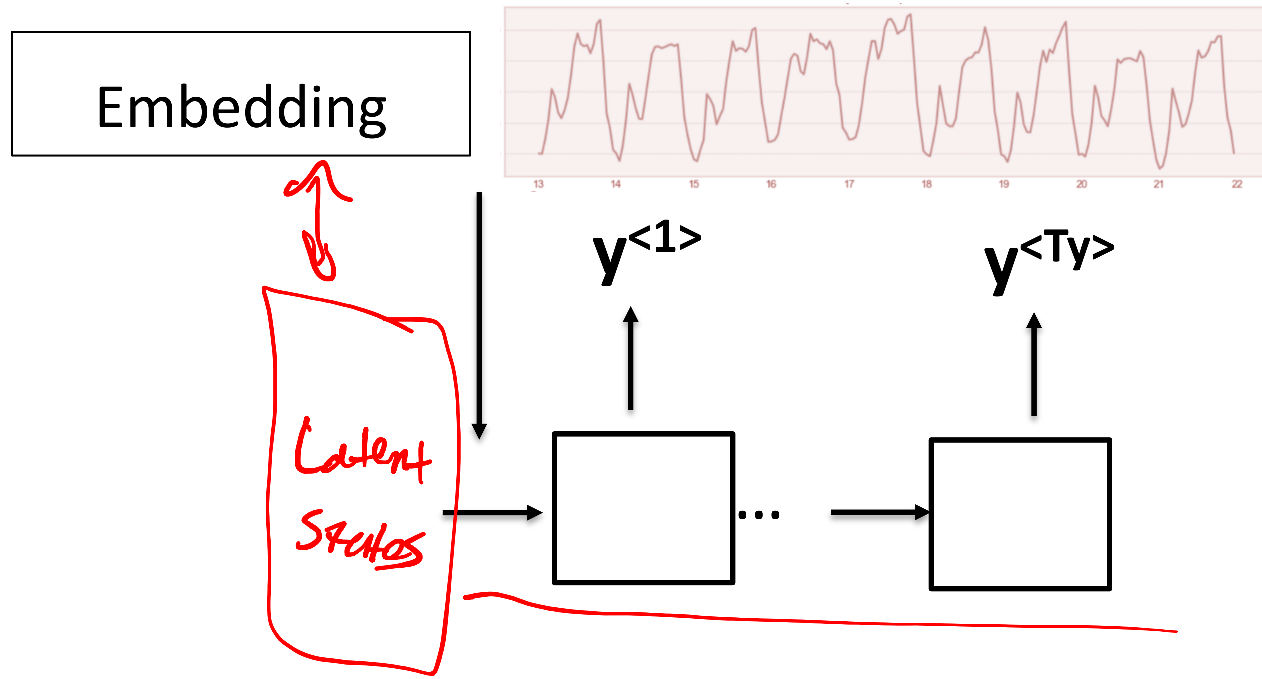


Reconstruction



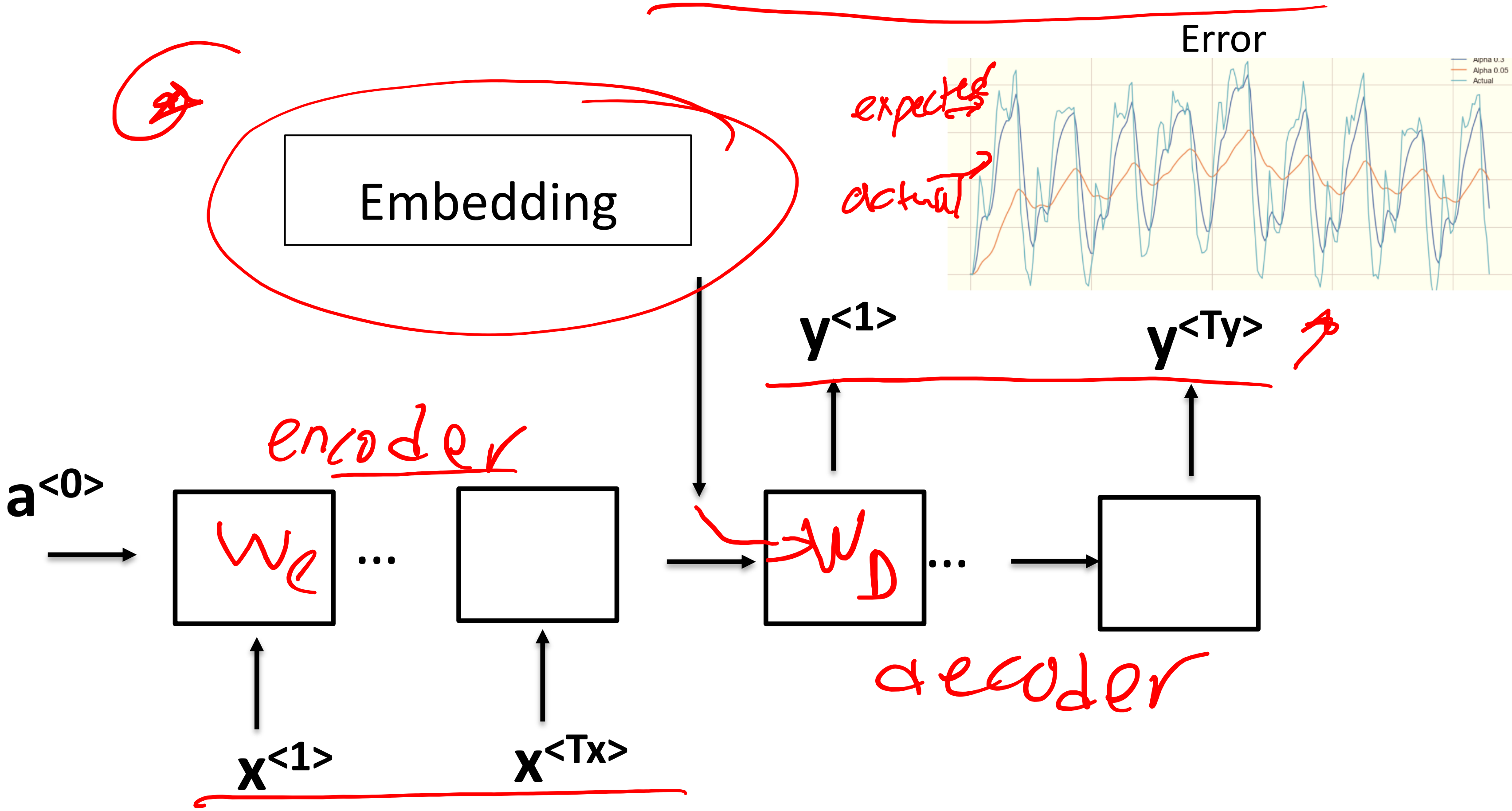
(Auto) Encoder

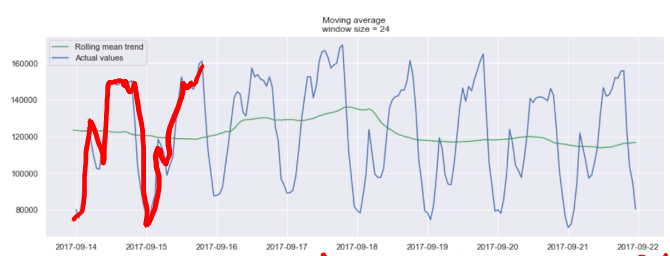
Reconstruction



(Auto) Decoder

Signal Validity (Anomaly Detection)





input = output