Deep learning Recap:

Neural Networks

- How to learn? (SGD)
- Nonlinearities (ReLU)
- Generalize well
- Image processing w/ conv layers

Text processing? Sequence processing
- Can be any length
- really not permutation-invariant
- words aren't vectors (know how to solve)

How to pass a sequence into a neural network?
- padding to max length ⇒ concat vectors
- ngram counts (usually intractible)
- recurrent neural network
- transformer network

RNN

DFA

letter in alphabet

current state

$$T(X, S) = S'$$

next state

current state $\mathbb{R}^d$

RNN:

$$T(X, S; W) = S'$$

input seq element as a vector $\mathbb{R}^m$

state

next state

$$A(S; U) \in \{-1, 1\} \in Y$$

function that maps from state to classification

$$h(x_1, x_2, \ldots, x_k) = A(T(T(\ldots T(x_1, S_0; W) \ldots)))$$

$$S_1 = T(x_1, S_0; W)$$

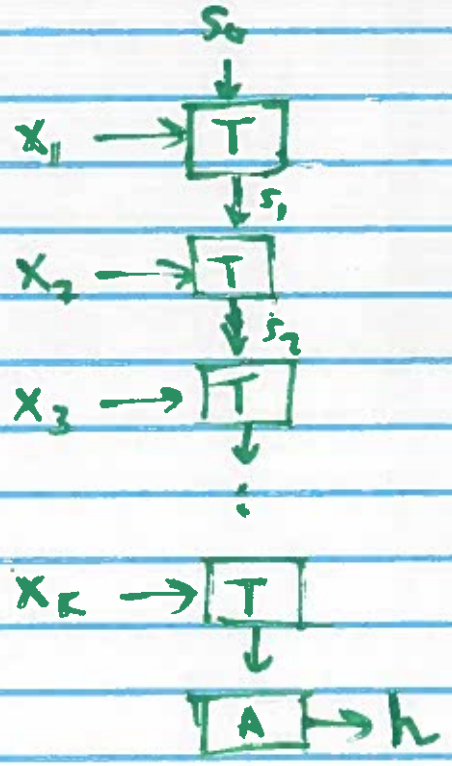
$$S_2 = T(x_2, S_1; W)$$

$$\vdots$$

$$S_i = T(x_i, S_{i-1}; W)$$

$$\vdots$$

$$h(x_1, \ldots x_k) = A(S_k; U)$$

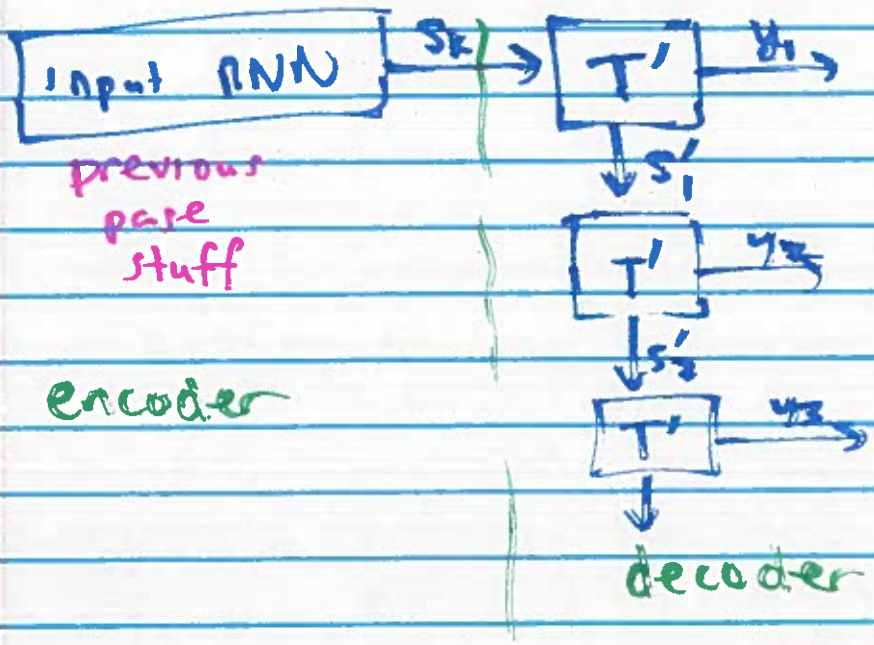


params: (W, U, S_0)

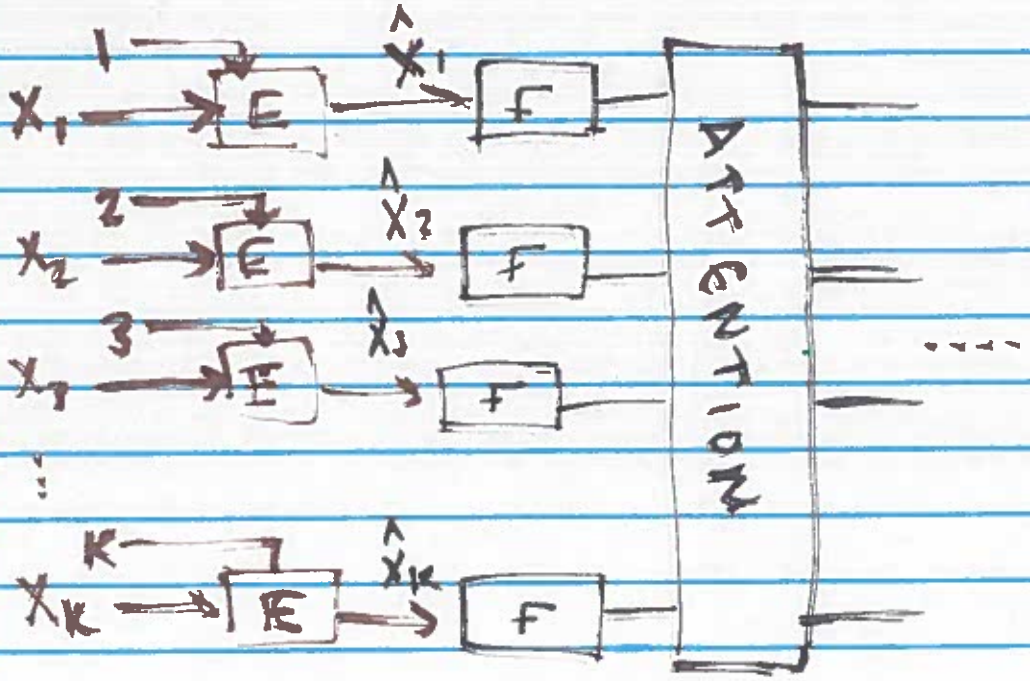
(e.g. LSTM)

RNNs for "Sequence-to-sequence"

Output-a-sequence



Another way: "Transformer Net" style



Privacy

I have dataset D.
Parts of it are private.

Goal: don't leak the training set!
a lot of learning algs aren't private

Most popular solution: differential privacy.

Adversarial Examples

My enemy control a small fraction of training data. Can they disrupt learning?

I have a network. Can adversary design an example that fools it?

Fairness

ML models are acting in the world.

They should be fair.