Sequences

Assignment

- Due Apr 4

- Any questions?

Recap

- Deep Learning Modules
  - Fully Connected
  - CNN
  - RNN (LSTM)

Today

- LSTM Modeling Examples
- Case Studies and Thought Experiments
- Creative combinations

LSTM Overview

- A differentiable memory cell
- Solves™ vanishing gradient
- Typically applied to sequence problems (not really)

Sequence

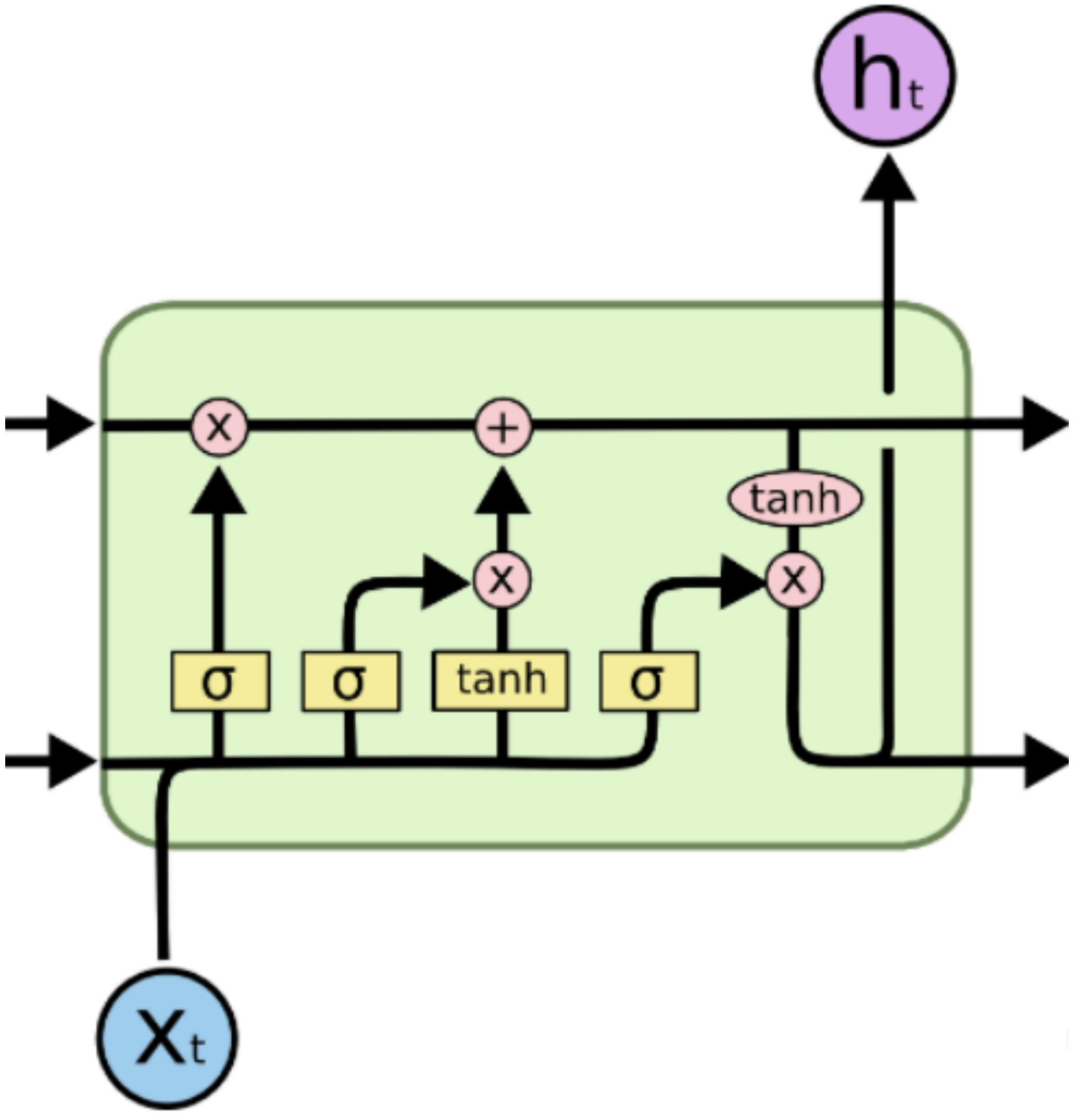
$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$

$$x^t$$

- Fixed length vector

LSTM

- Cell state
- Consume input + hidden state
- Update cell state
- Return output + new hidden state


h_t
x
+
tanh
x
x
σ
σ
tanh
σ
X_t

A Good Heuristic

$$f(x, h^{t-1}; C^{t-1})$$

$$h^t, o^t$$

So

$$f(x, h^{t-1}; C^{t-1})$$

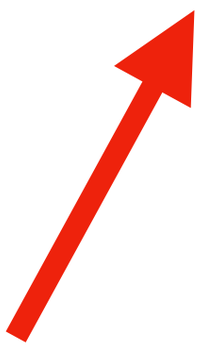
Input

Previous Hidden state

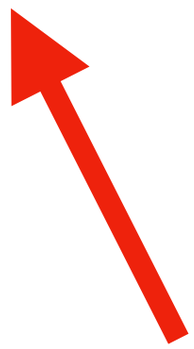
Cell state

And

$$h^t, o^t$$



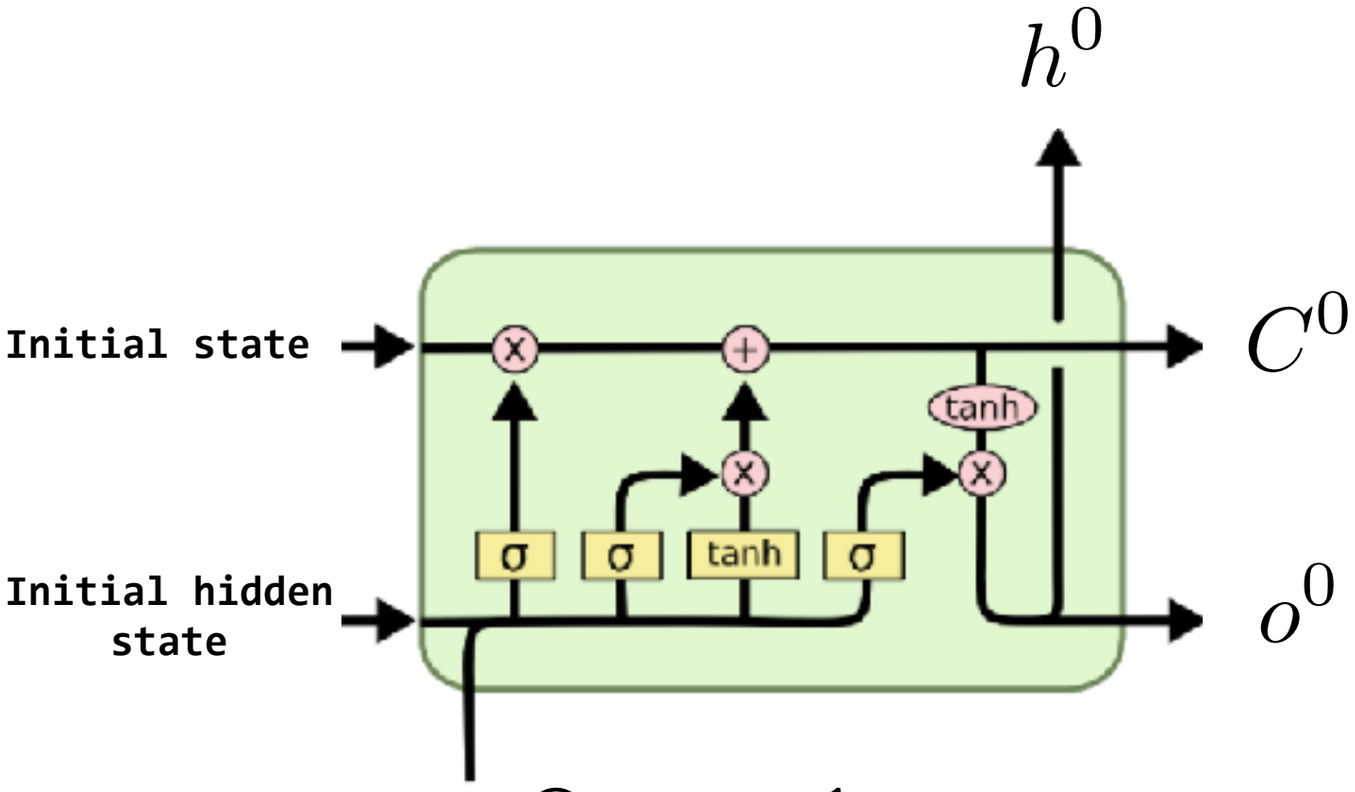
New hidden state



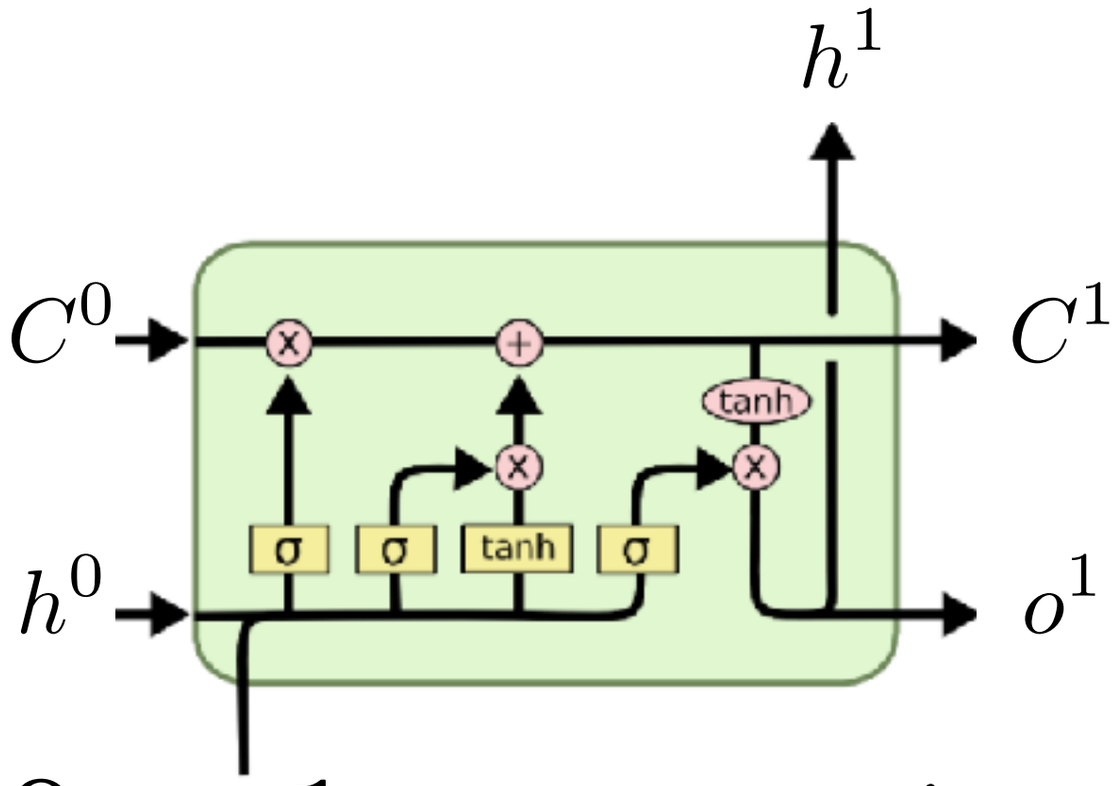
Output

And

- Consume current position in sequence
- Update cell state (internal)
- Return new hidden state
- Return new output



$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$



$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$

And So On

At The End

$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$

$$< h^0, h^1, \ldots, h^t, \ldots, h^n >$$

And

$$< o^0, o^1, \ldots, o^t, \ldots, o^n >$$

Ok So What?

Handwriting

Input

Input

- A sequence of feature vectors
  - For each column of input
    - Row-id of black pixel
    - Position of top black pixel
    - . . .
    - . . .

Input

$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$

Output

- Handwriting recognition
  - Handwritten text -> character sequence

Output

- Character sequence

- One hot character vectors

Thus

$$< o^0, o^1, \ldots, o^t, \ldots, o^n >$$

One hot vectors

So

- Basically one classification problem per input
- Include special ϵ symbol in your character set

Consider

Problem

- Sequence input

- Need to predict just 1 output (say yes/no) for the whole sequence

Example

- Sequence of test grade scores
- Predict if the student will pass course or fail

Ideas?

Dataset

- Input:
  - Grades sequence
- To predict:
  - P / F
  - So binary classification

LSTM Setup

- Input:

$$< x^0, x^1, \ldots, x^t, \ldots, x^n >$$

LSTM Setup

- Output:

$$< o^0, o^1, \ldots, o^t, \ldots, o^n >$$

$$< o^0, o^1, \ldots, o^t, \ldots, o^n >$$

Ignore

So

$$O^n$$

This Is

- The final output after the whole sequence is consumed

Just use this

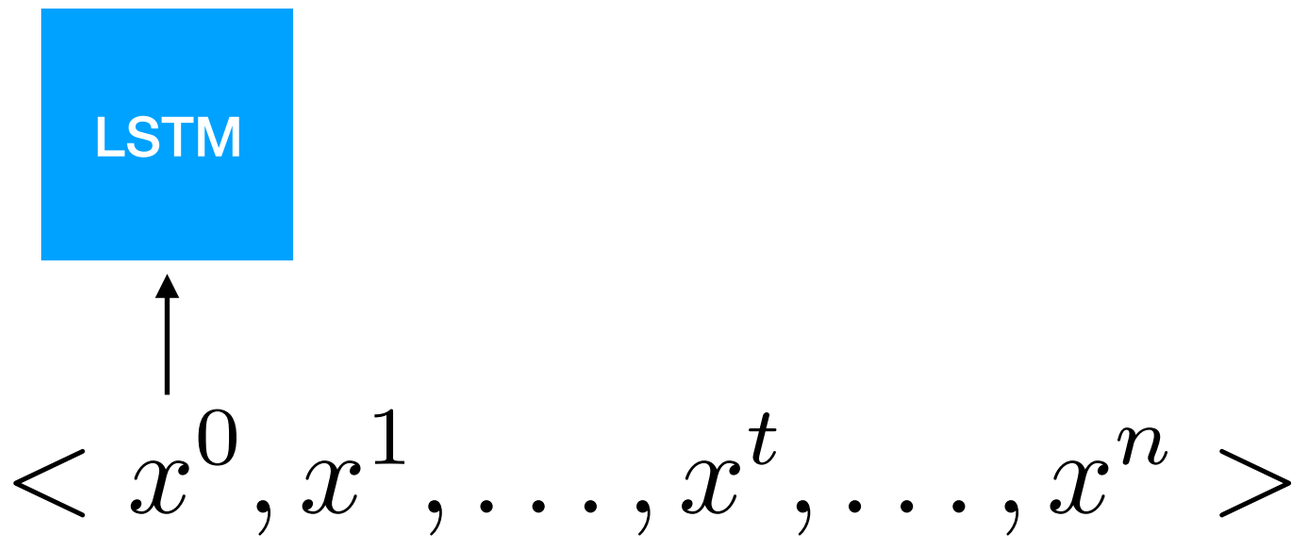
Using O^n

- It is a vector
- Need to use it for P / F prediction

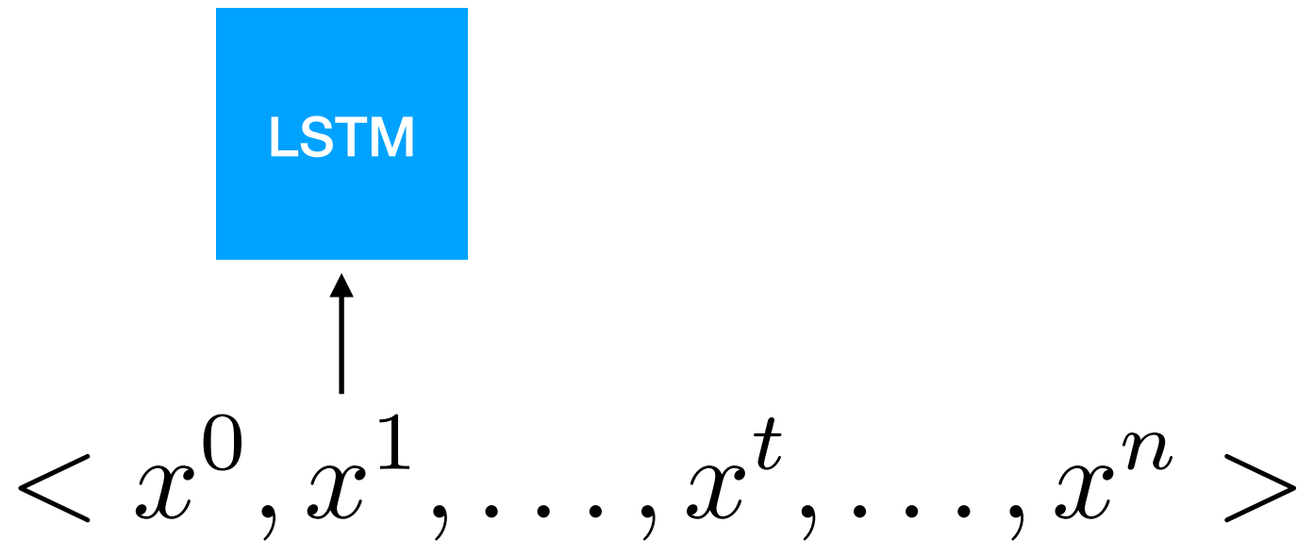
MLP

- Fully connected layers operate on O^n

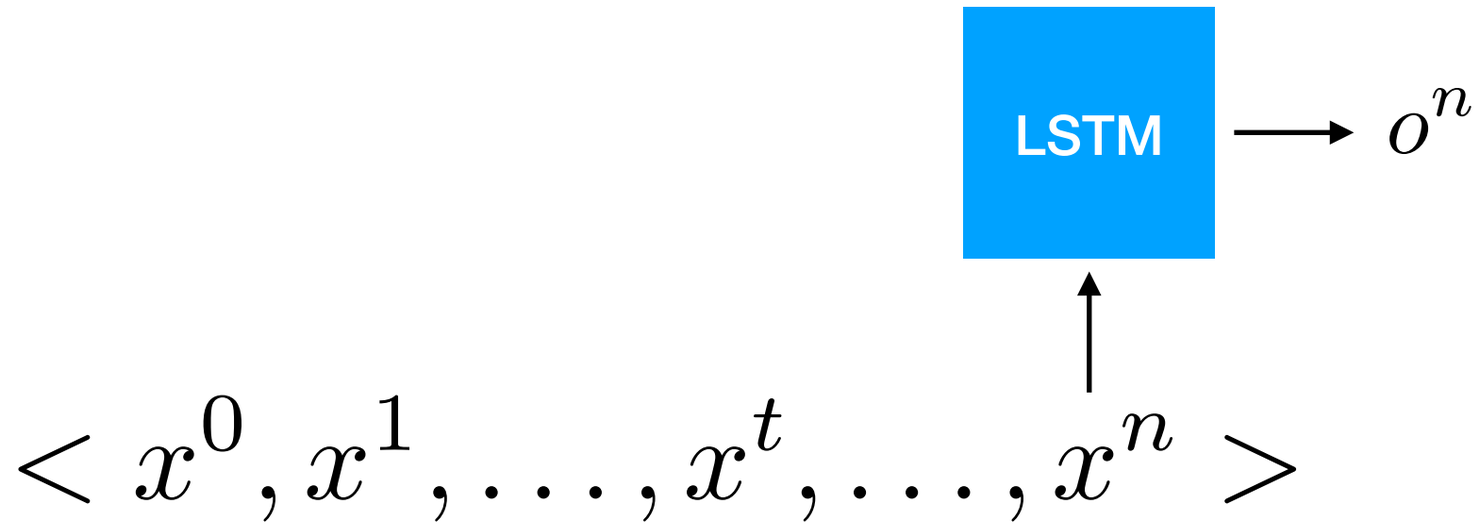
The Big Picture



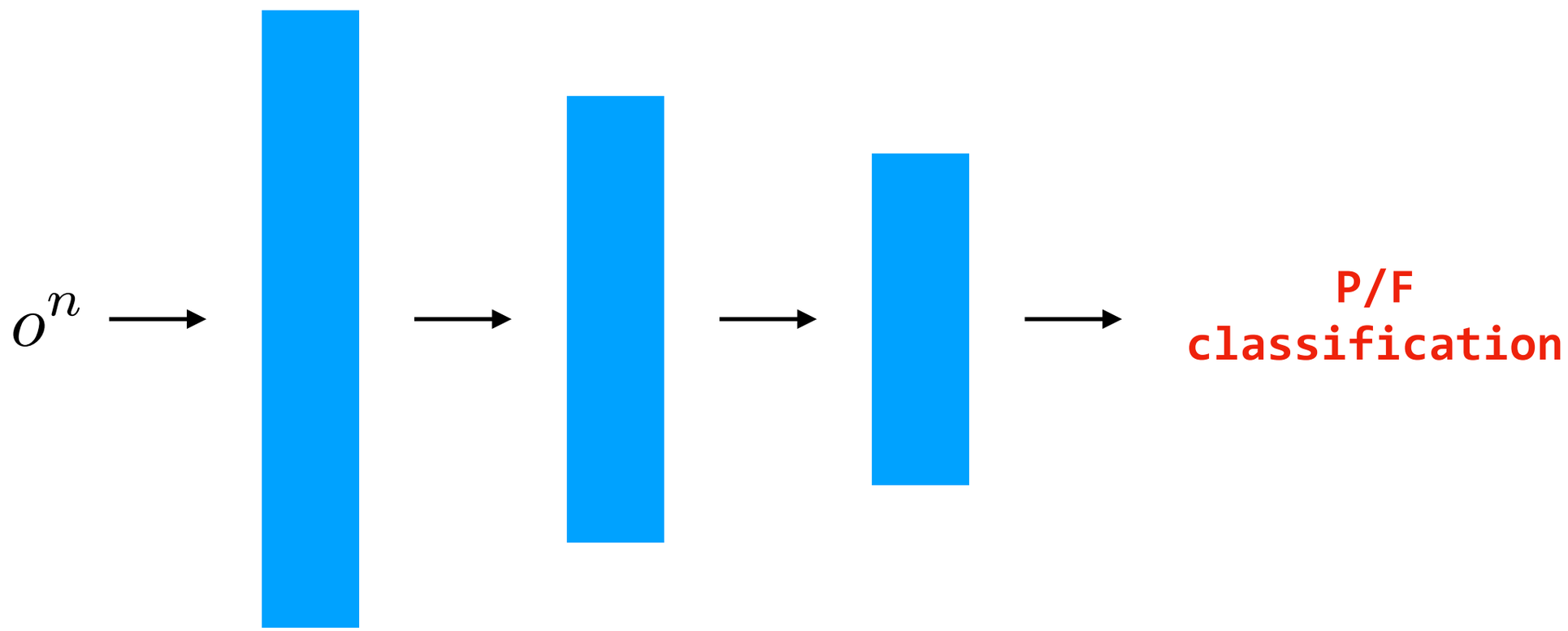
The Big Picture



The Big Picture



The Big Picture



Same Stuff

- Convolutional backbone
- Input -> Representation
- Representation fed into MLP for prediction

LSTM

- Sequence input
- Fed into LSTM
- Get output(s)
- Feed them into other things to do clever stuff

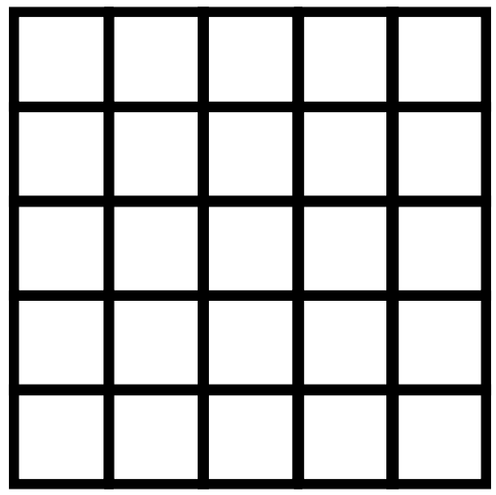
MODULAR THINKING!

MNIST Example

Input

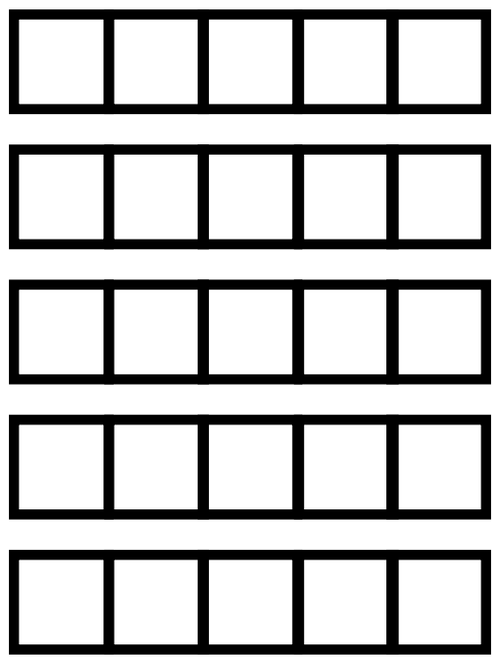
- 28 x 28 image

Image



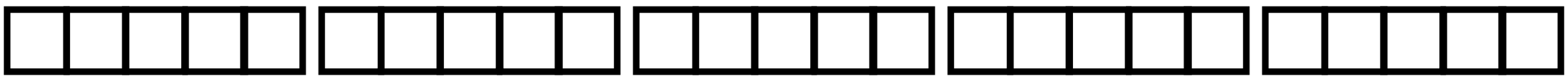
5 x 5 Image

Image



Rows of pixels

Image



This is

- A sequence of pixel row vectors

MNIST

- Image -> Digit

LSTM Setup

- Sequence of pixel rows as input
- Take final output
- 10-category classification

Code

Acknowledgements

- LSTM cell diagram from Chris Olah’s fantastic blog post: http://colah.github.io/posts/2015-08-Understanding-LSTMs/