HW4P1 Bootcamp

Additional Resources S23

Language Model

- Just read the write up :) Its explained very well
- These slides will be a brief overview of the concepts for your understanding alone. You will have to add in extra steps according to the instructions given in the starter notebook (Eg. adding sos, eos)

Language Model

I took a block of ice from the refrigerator and watched it for a while. I was watching "Frozen"

The reason why restaurants have starters before the main course is to eat

Language Model

I took a block of ice from the refrigerator and watched it for a while. I was watching "Frozen"

The reason why restaurants have starters before the main course is to eat.
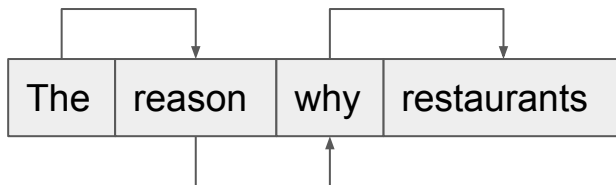
Consider this amazing joke

The	reason	why	restaurants

Language Model

I took a block of ice from the refrigerator and watched it for a while. I was watching "Frozen"

The reason why restaurants have starters before the main course is to eat.

Consider this amazing joke

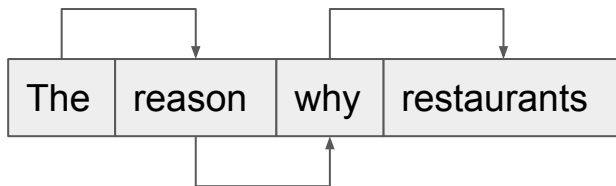| The	reason	why	restaurants

Language Model

I took a block of ice from the refrigerator and watched it for a while. I was watching "Frozen"

The reason why restaurants have starters before the main course is to eat.

Consider this amazing joke

The	reason	why	restaurants

The -> *reason*. Given these 2 words, the next word should be *why* and so on if you are modelling bad jokes

Dataset

- You have a list of articles. Every article is of different length (579,)

Dataset

- You have a list of articles. Every article is of different length (579,)
- Concatenate

Input:

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t

Dataset

- You have a list of articles. Every article is of different length (579,)
- Concatenate
- Create output

Output:

| b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t |

<- so any arrows ->

Input:

| a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t |
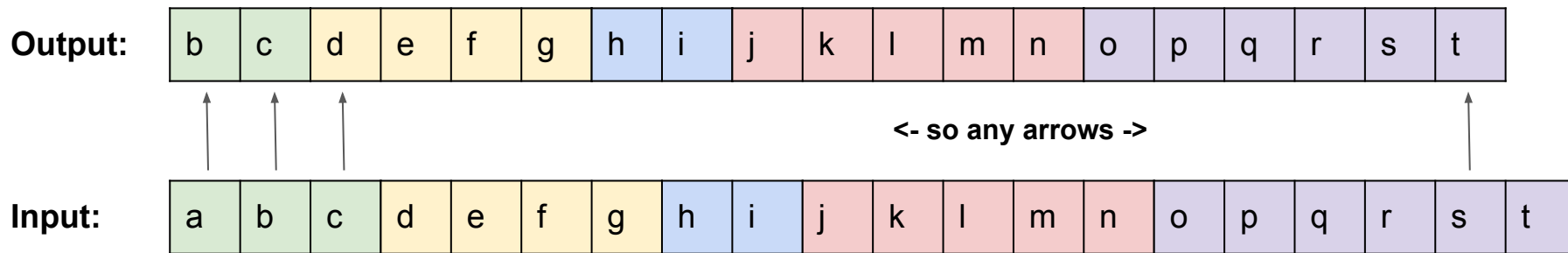
Don't need to worry about the last word in every article since the articles are pretty long

Dataset: Batching

- If your batch size is 2 and your sequence length is 4, how many words do you have in a batch?

Dataset: Batching

- If your batch size is 2 and your sequence length is 4, how many words do you have in a batch?
- You have batch_size * seq_len = 8 words

Sample 1 -> a	b	c	d
Sample 2 -> p	q	r	s

What is the shape of the output?

Dataset: Batching

- If this is the input. What is the output? Can you predict?

Sample 1 ->	a	b	c	d
Sample 2 ->	p	q	r	s

Dataset: Batching

- If this is the input. What is the output? Can you predict?

Sample 1 ->	a	b	c	d
Sample 2 ->	p	q	r	s

- This is the output.

Sample 1 ->	b	c	d	e
Sample 2 ->	q	r	s	t

Dataset: Batching

- If your batch size is 2 and your sequence length is 4, how many words do you have in a batch?
- You have batch_size * seq_len = 8 words
- You need to yield batches like these from the whole dataset

Output:

| b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t |

Input:

| a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t |

Batch 1   Batch 2   ...   Batch N
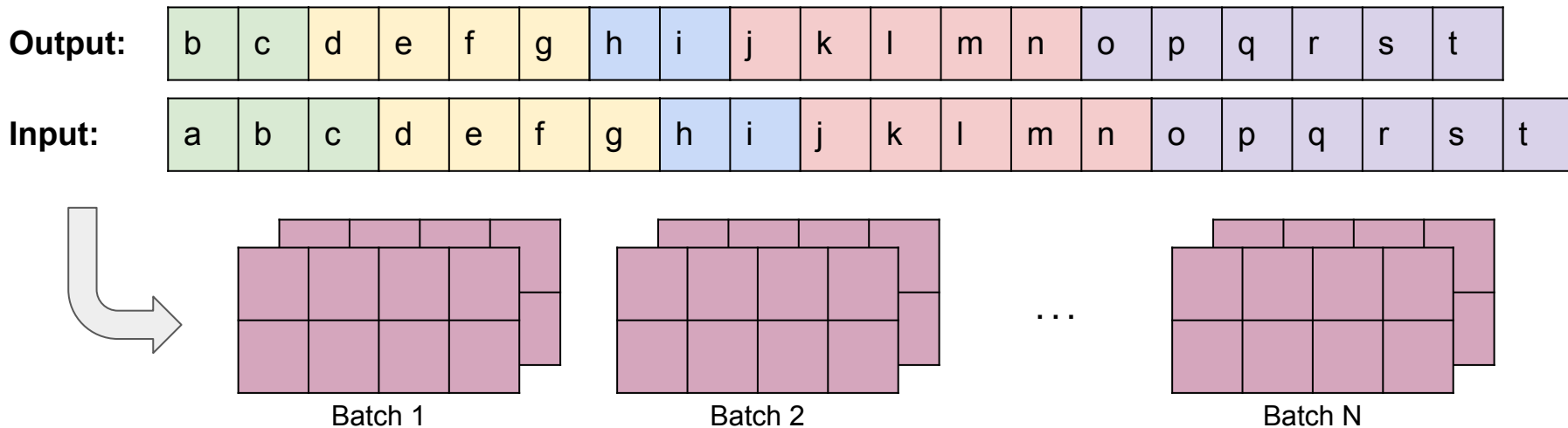
Dataset: Batching

- Gonna leave reshaping to return batches up to you
- Note: If you are using variable sequence length then it might be different

New modules for this HW

- For all these, please read the pytorch docs first :')
  - nn.Embedding (Explained in hw3p2 bootcamp actually)
  - LSTMCell - Normal LSTM but just for 1 timestep
- All these will visit you in HW4P2 as well :-0
- Things to try (some of these might help in p2 as well)
  - Dropout for LSTMCells
  - Weight tying
  - Embedding dropout
  - Many others given in the paper linked in the writeup

Problems: Prediction

$Y_8 = P(t_8|\text{<sos>FOUR S})$

$P(\text{<eos>})$
$P(A)$
$P(B)$
$P(C)$ ← $P(C|\text{<sos>FOUR S})$
$P(Z)$

<sos> F O U R ☐ S
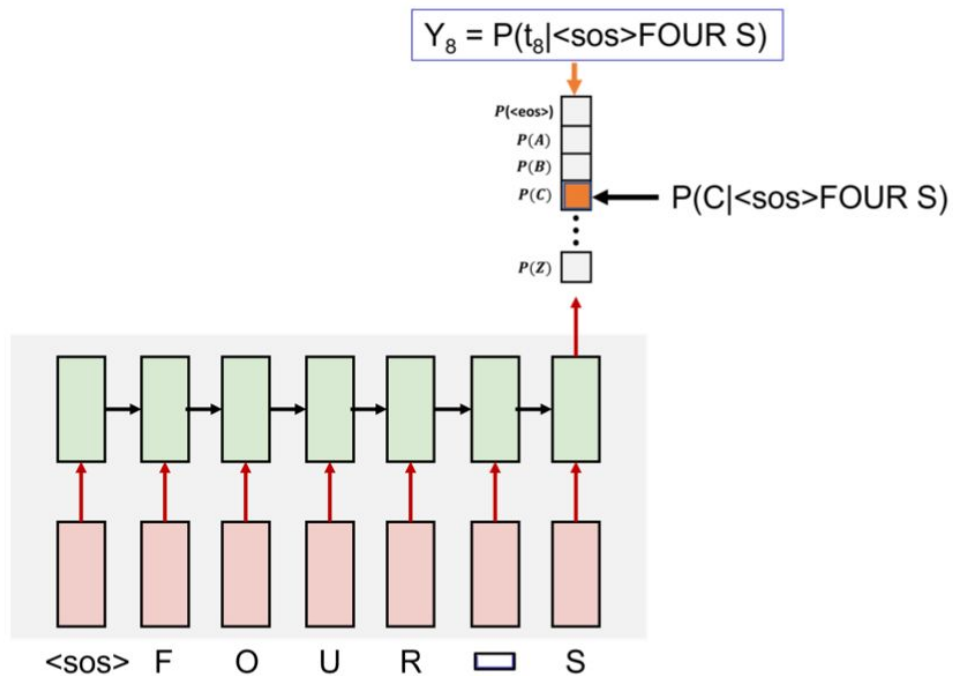
Problems: Generation
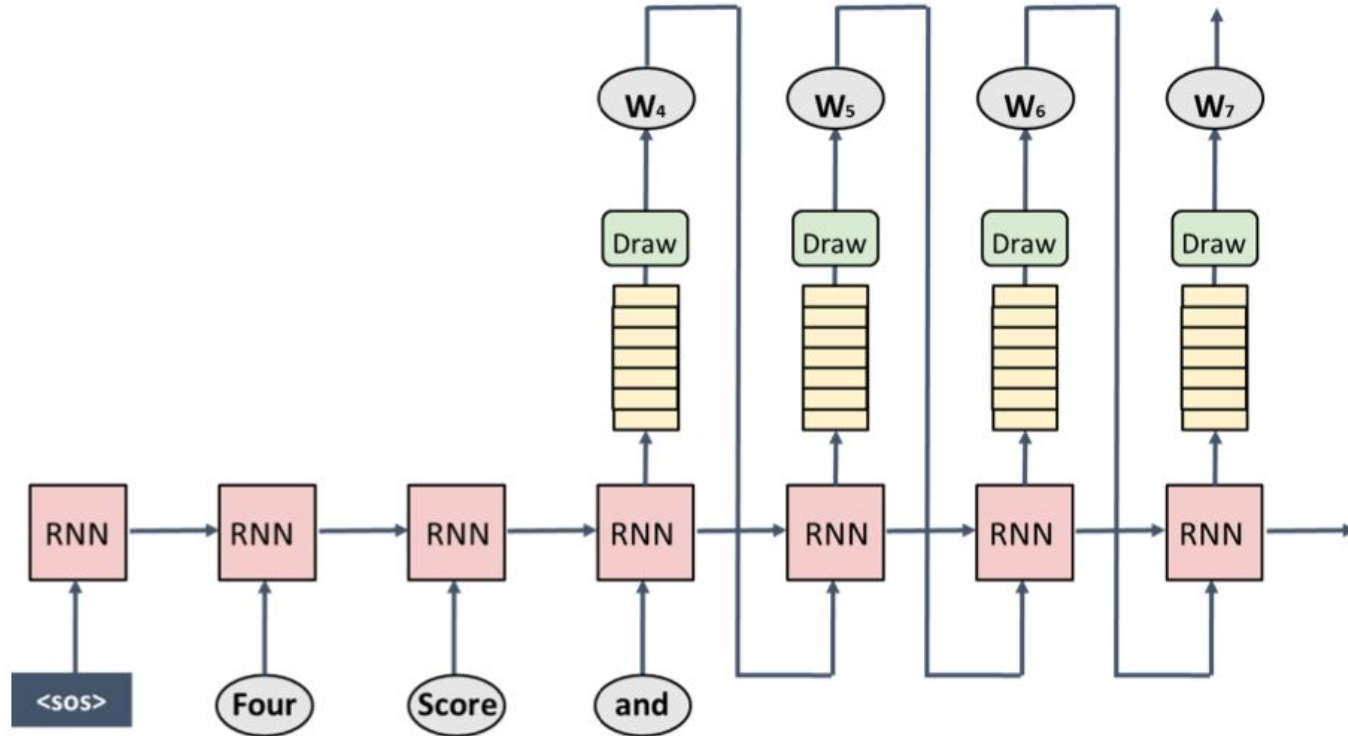
General tips

- The LM in this HW is equivalent to "Speller" in HW4P2. You can do ablations for the Speller in this HW (It won't be exactly similar but it's a good start). In addition, training in this HW only takes < 2min/epoch with FP16
- Write up and the starter notebook are pretty detailed for you to understand what needs to be done.