Language Modelling for Text Prediction

Devan Kuleindiren

Do you want to grab a _____

Do you want to grab a

drink	0.327
coffee	0.211
bite	0.190
spot	0.084
⋮	⋮

N-gram

N-gram

$$P(w_i|w_{i-n+1} \dots w_{i-1}) = \frac{count(w_{i-n+1} \dots w_{i-1} w_i)}{\sum_w count(w_{i-n+1} \dots w_{i-1} w)}$$

N-gram

$$P(\text{sat}|\text{the cat}) = \frac{count(\text{the cat sat})}{\sum_w count(\text{the cat } w)}$$

N-gram

$$P(\text{sat}|\text{the cat}) = \frac{count(\text{the cat sat})}{\sum_w count(\text{the cat } w)}$$

Backoff

$$P(\text{sat}|\cancel{\text{the}}\ \text{cat})$$

Recurse on an $(n-1)$-gram model.

N-gram

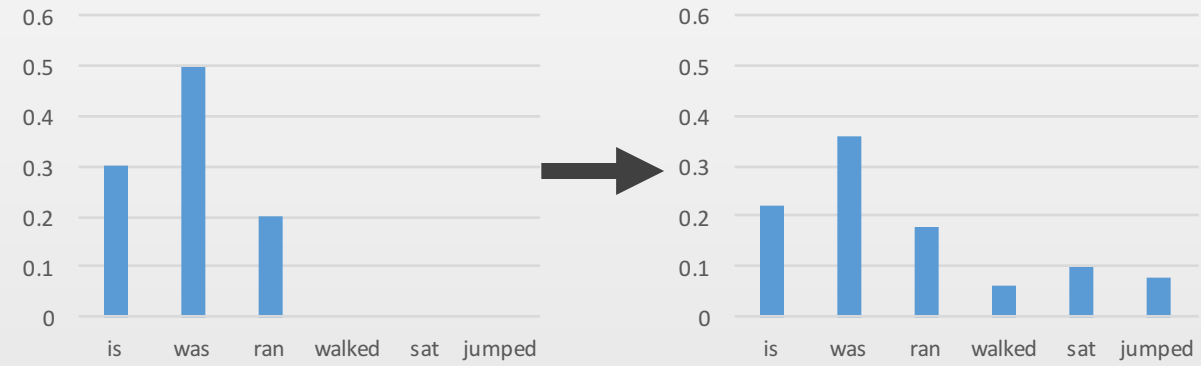
$$P(\text{sat}|\text{the cat}) = \frac{count(\text{the cat sat})}{\sum_w count(\text{the cat } w)}$$

Backoff

$$P(\text{sat}|\cancel{\text{the}}\ \text{cat})$$

Recurse on an $(n-1)$-gram model.

Smoothing

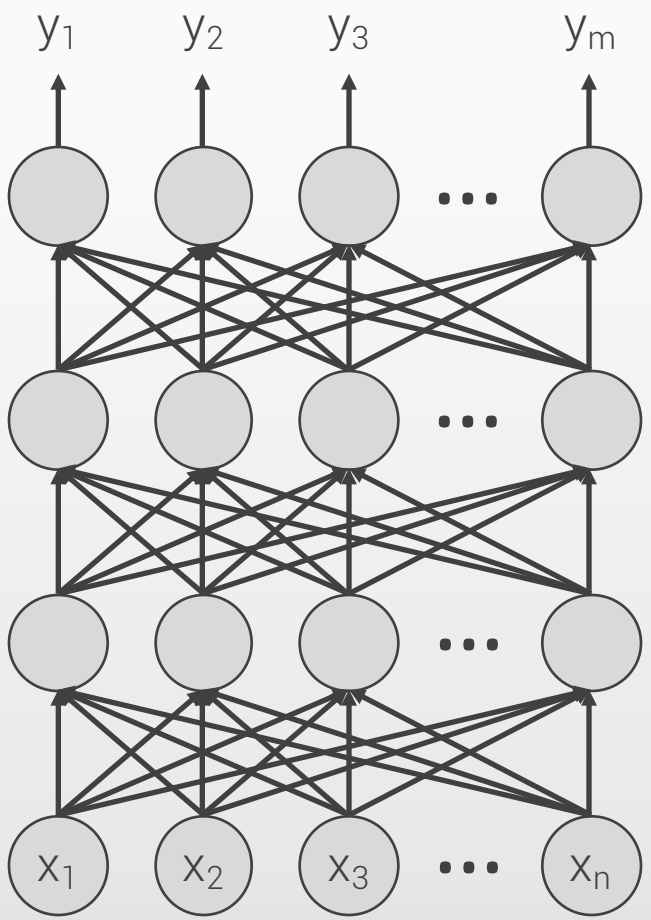


Some methods:

- Add one
- Katz
- Absolute discounting
- Kneser-Ney
- Modified Kneser-Ney

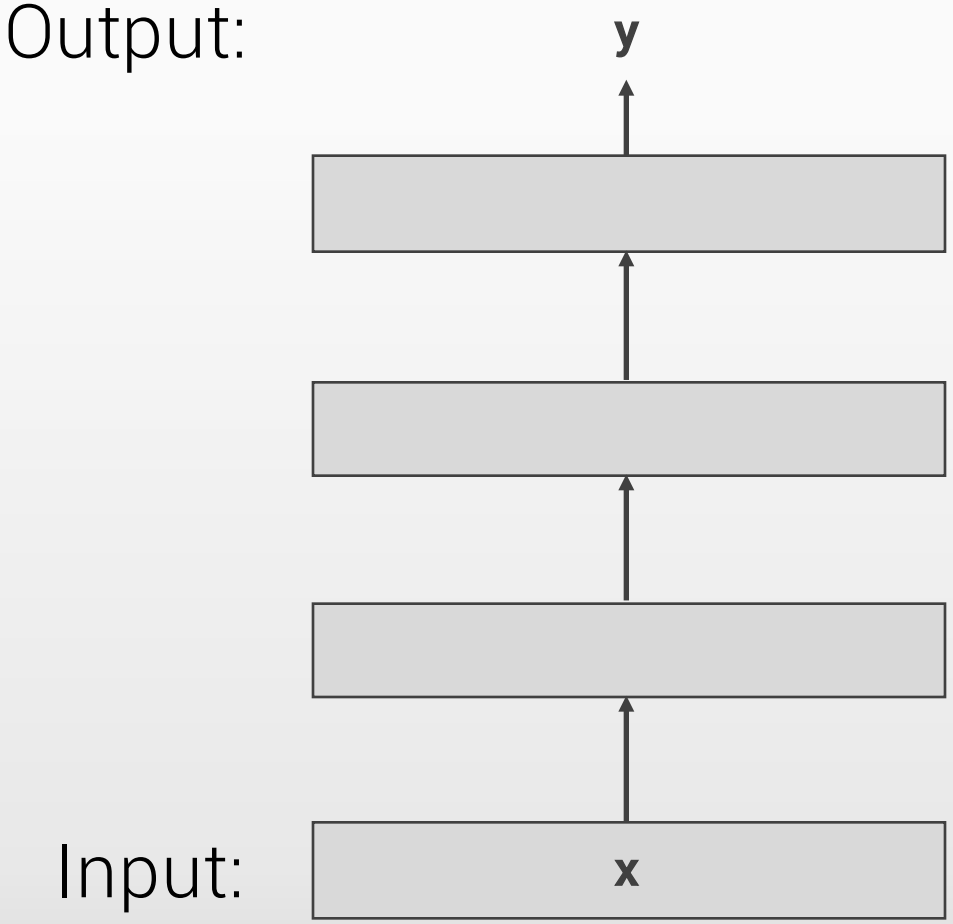
Neural Networks

Output:

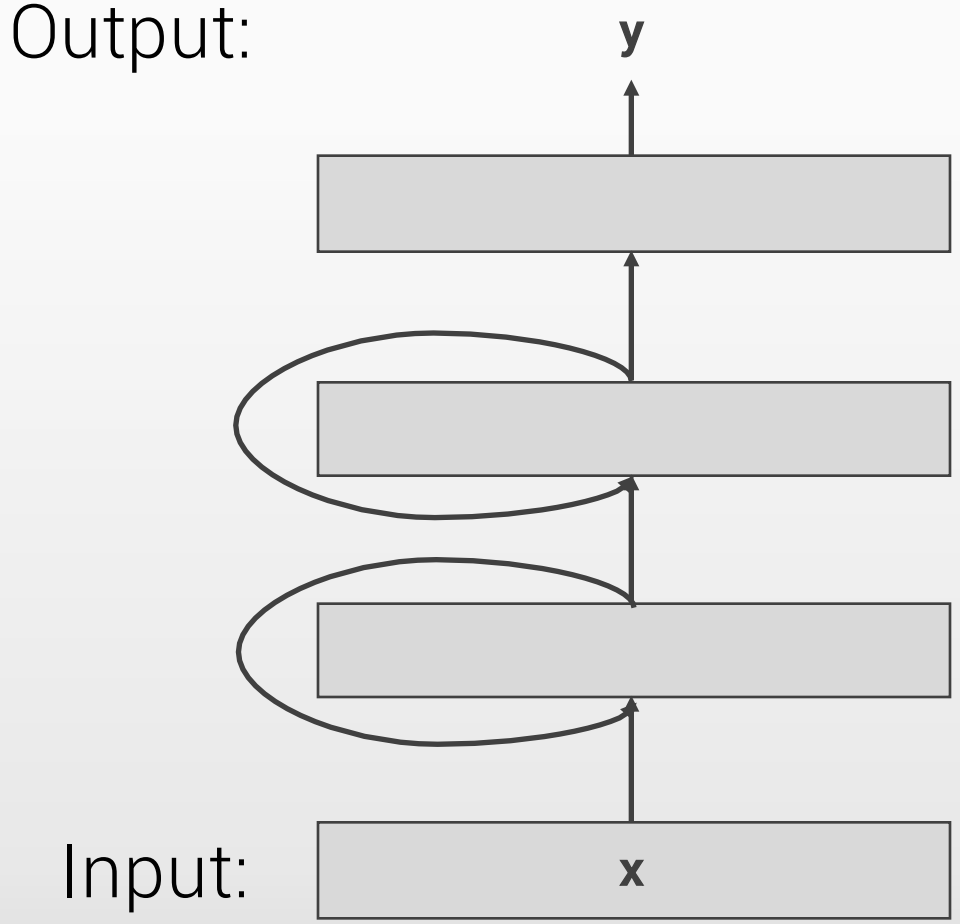


Input:

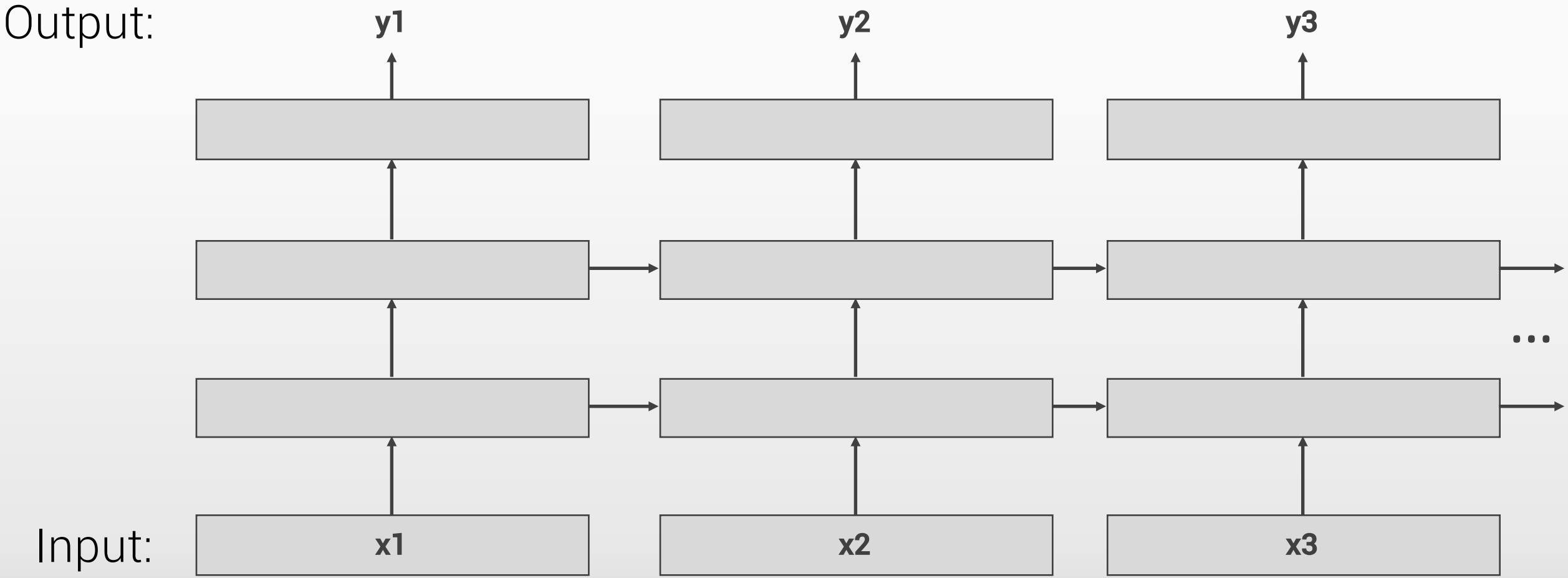
Neural Networks



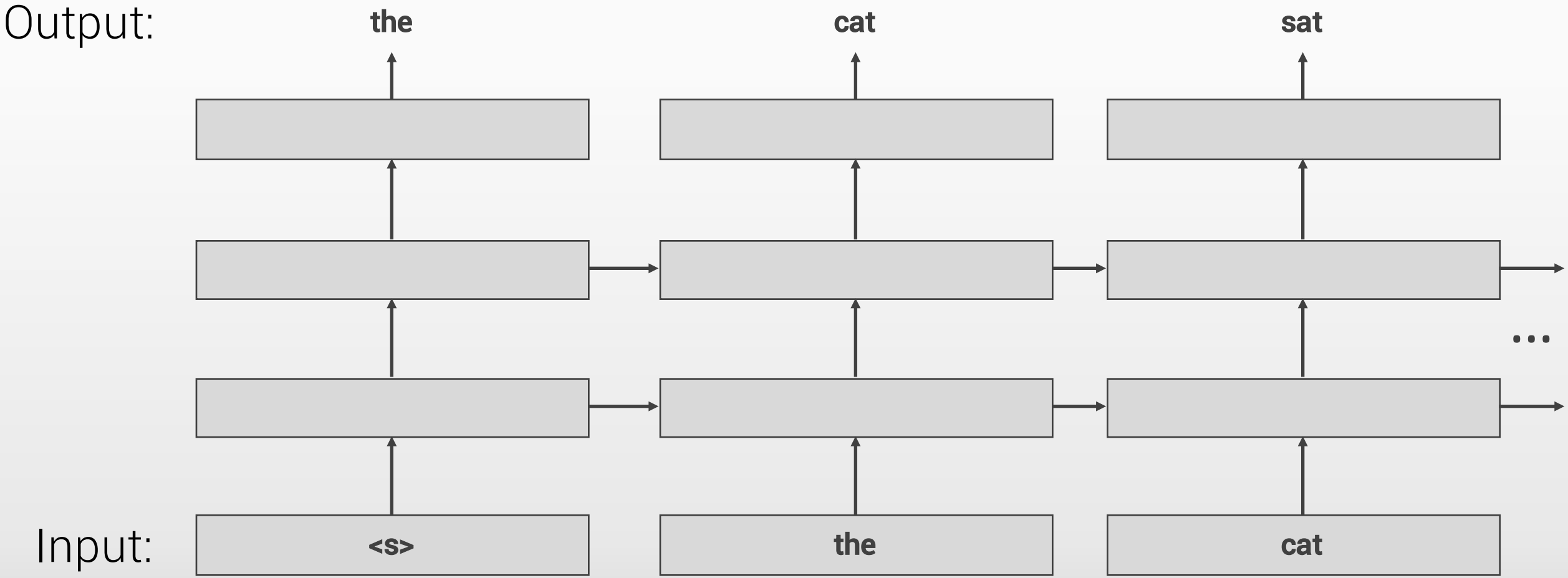
Recurrent Neural Networks



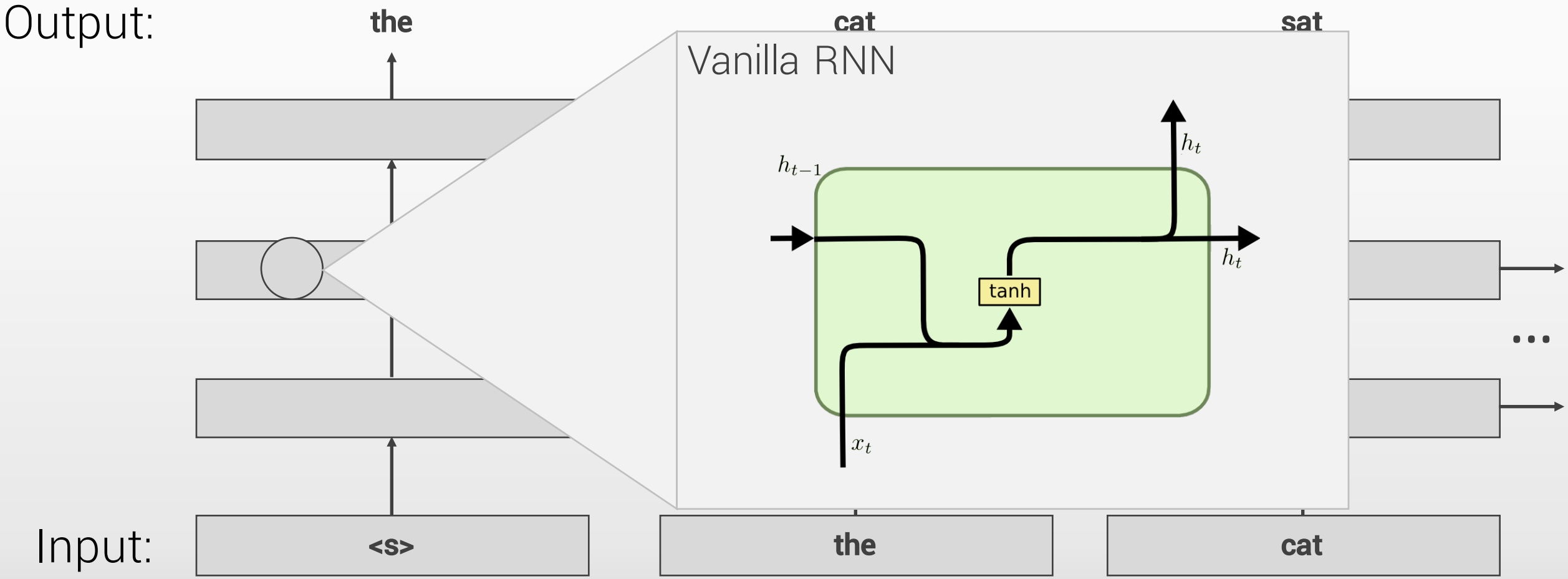
Recurrent Neural Networks



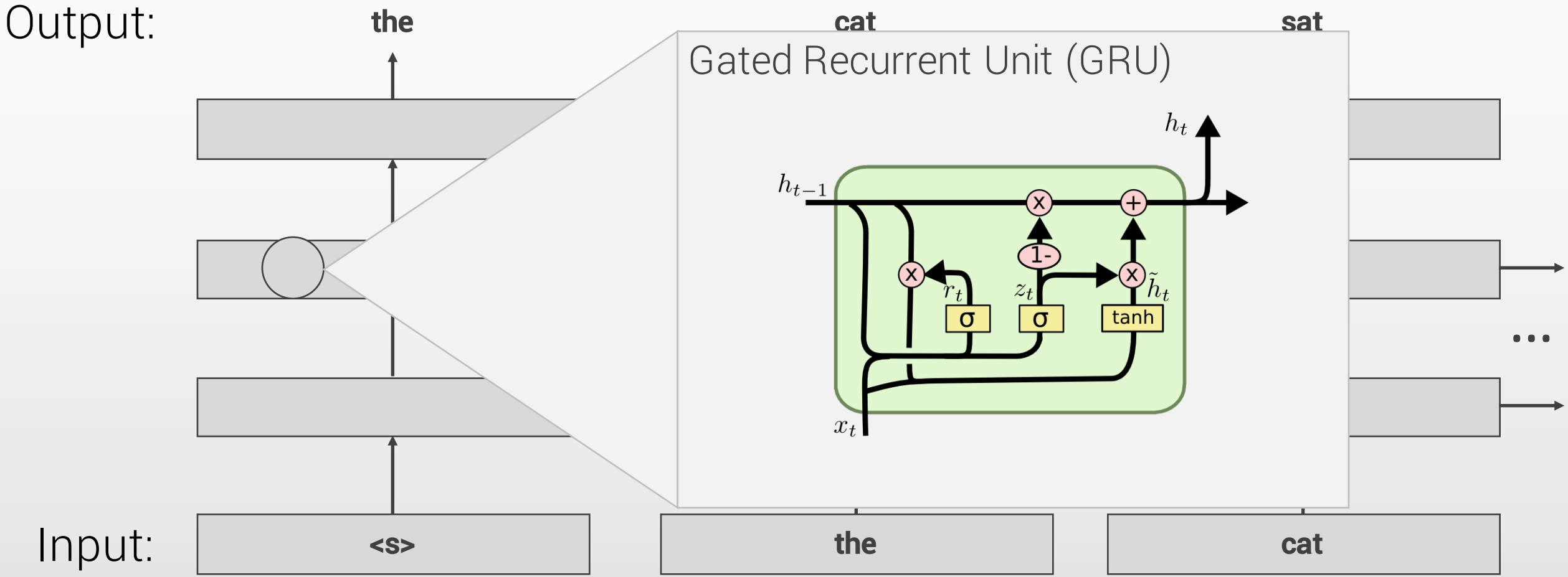
Recurrent Neural Networks



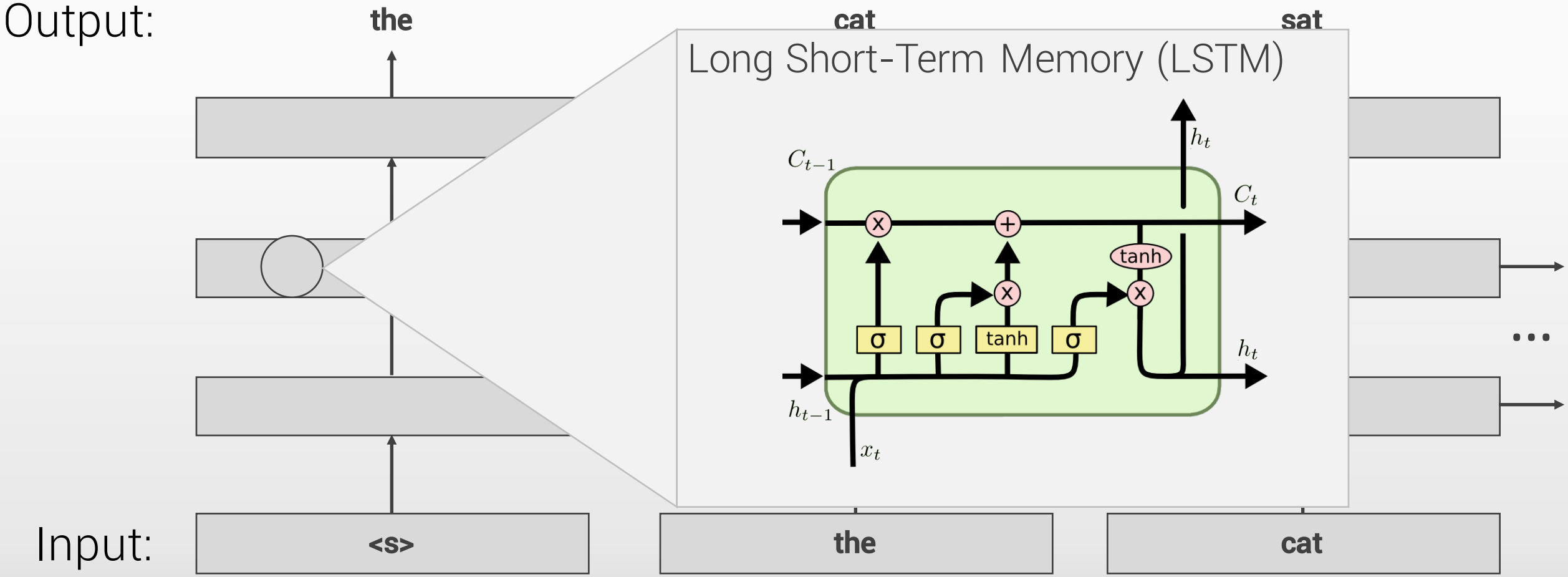
Recurrent Neural Networks



Recurrent Neural Networks

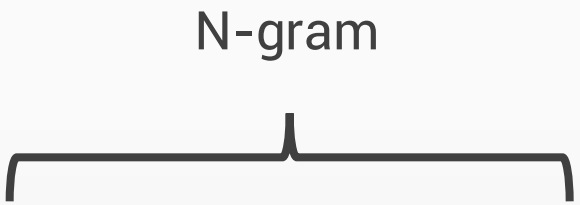


Recurrent Neural Networks



What have I built?

Progress



+ Add one
+ Katz
+ Absolute discounting
+ Kneser-Ney
+ Modified Kneser-Ney

Progress

N-gram

+ Add one
+ Katz
+ Absolute discounting
+ Kneser-Ney
+ Modified Kneser-Ney

RNN

Vanilla RNN

Gated Recurrent Unit

Long Short-Term Memory

Progress

N-gram

+ Add one
+ Katz
+ Absolute discounting
+ Kneser-Ney
+ Modified Kneser-Ney

RNN

Vanilla RNN

Gated Recurrent Unit

Long Short-Term Memory

Benchmark

- Perplexity
- Average keys saved
- Guessing entropy
- Physical memory usage
- Prediction speed

Progress

N-gram

+ Add one
+ Katz
+ Absolute discounting
+ Kneser-Ney
+ Modified Kneser-Ney

RNN

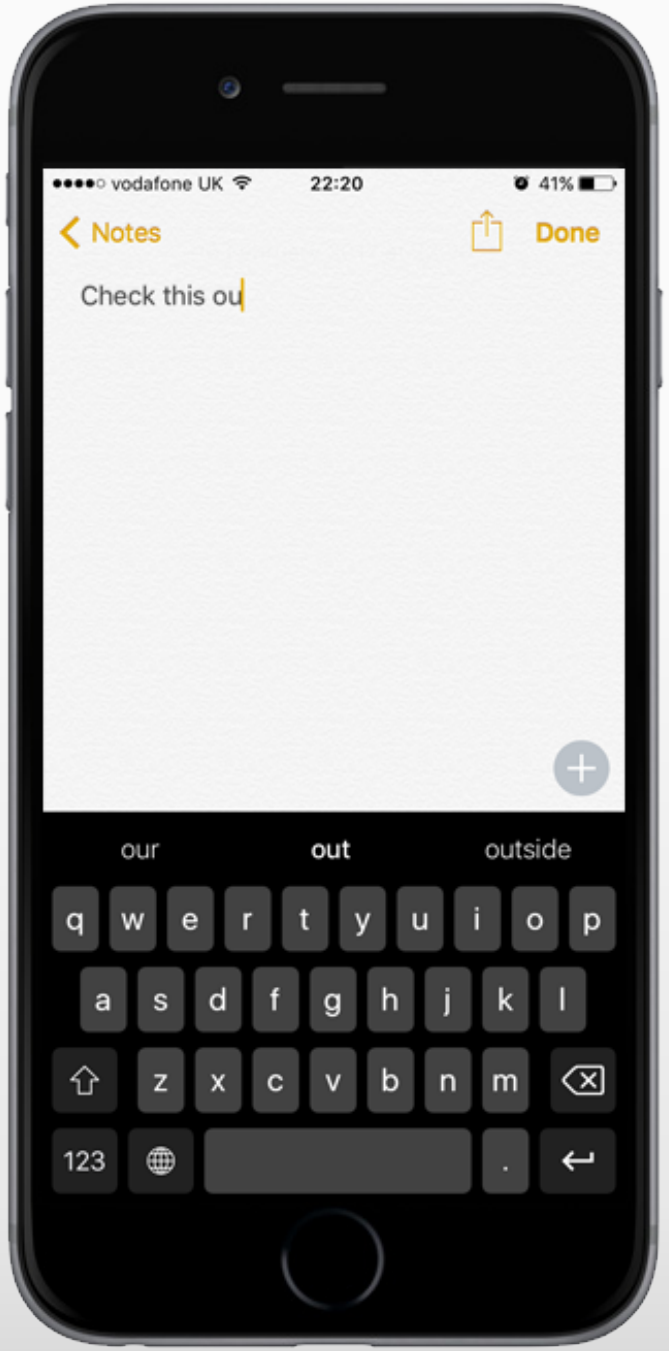
Vanilla RNN

Gated Recurrent Unit

Long Short-Term Memory

Benchmark

- Perplexity
- Average keys saved
- Guessing entropy
- Physical memory usage
- Prediction speed



What's left?

What's Left

The weather is nice

What's Left

The whether is ____

3-gram + Kneser-Ney

it isn't the only way to get a piece of the company 's stock exchange composite trading on the new york stock exchange composite trading on the ...

LSTM

it is a way to get out of the company 's new strategy.

3-gram + Kneser-Ney

the meaning of life is threatened to veto the bill.

LSTM

the meaning of life is that the new york times isn't the only way to be the first time to be a major part of the problem.

Thanks for listening!