Sequence to sequence models

Beam search

deeplearning.ai

Beam search algorithm

$B = 3$ (beam width)

Step 1

$$\begin{bmatrix} a \\ \vdots \\ in \\ \vdots \\ jane \\ \vdots \\ september \\ \vdots \\ zulu \end{bmatrix} \quad 10000$$

$\rightarrow P(y^{<1>} \mid x)$

$\hat{y}^{<1>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square$
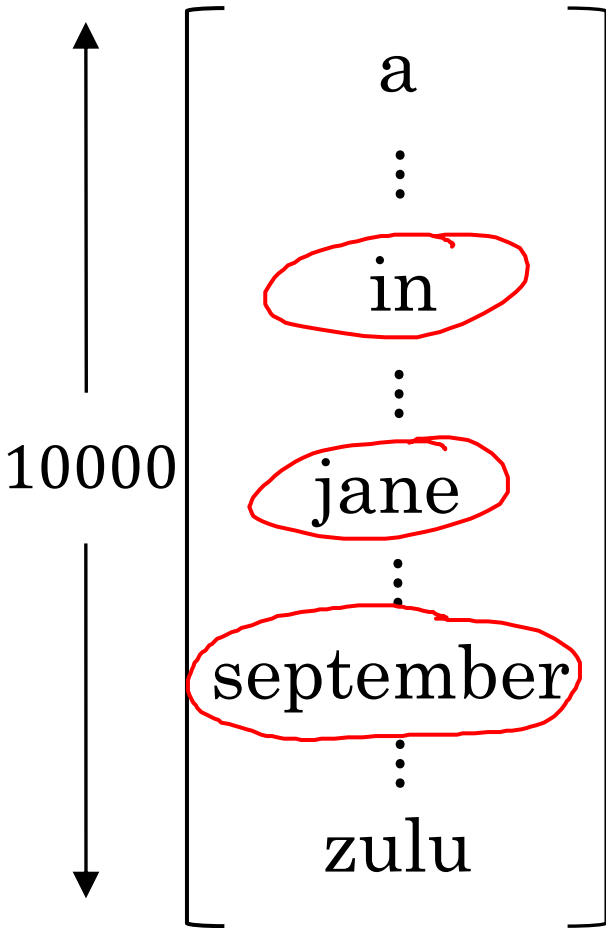
$x^{<1>} \qquad x^{<T_x>}$
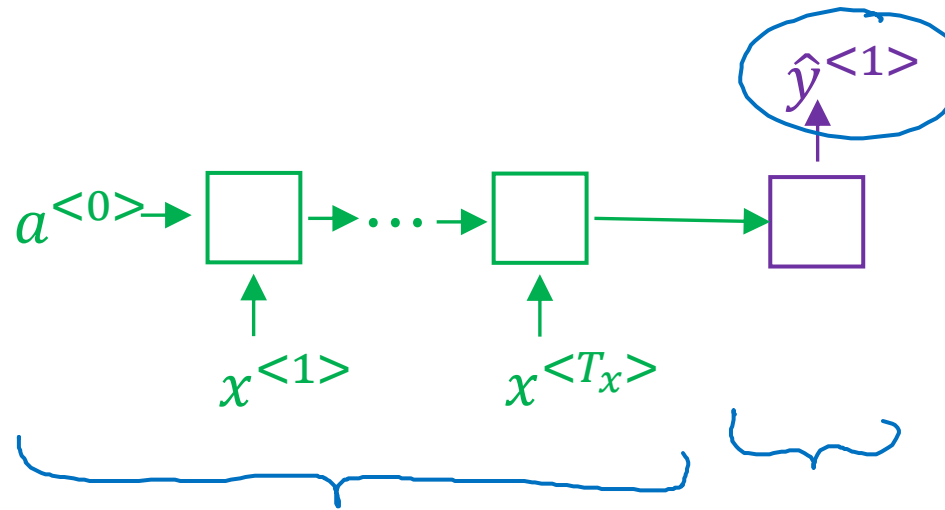
Try all possible word as the first word, and choose the top 3 which maximize the probability.

Andrew Ng

Beam search algorithm

$(\beta = 3)$

Step 1   Step 2

$$\begin{bmatrix} a \\ \vdots \\ in \\ \vdots \\ jane \\ \vdots \\ september \\ \vdots \\ zulu \end{bmatrix}$$

10000

$y^{<1>}, y^{<2>}$

a
aaron
September
visit
zulu

a
aaron
is
visit
zulu

10,000

a
zulu

$P(y^{<1>}, y^{<2>} | x) = P(y^{<1>} | x) P(y^{<2>} | x, y^{<1>})$

in
$\hat{y}^{<1>}$ $\hat{y}^{<2>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$

in

$P(y^{<2>} | x, "in")$

jane $\hat{y}^{<2>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$

$P(y^{<2>} | x, "jane")$

Try the second word based on the first word is given by step one, to maximize the probability

September $\hat{y}^{<2>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$

Andrew Ng
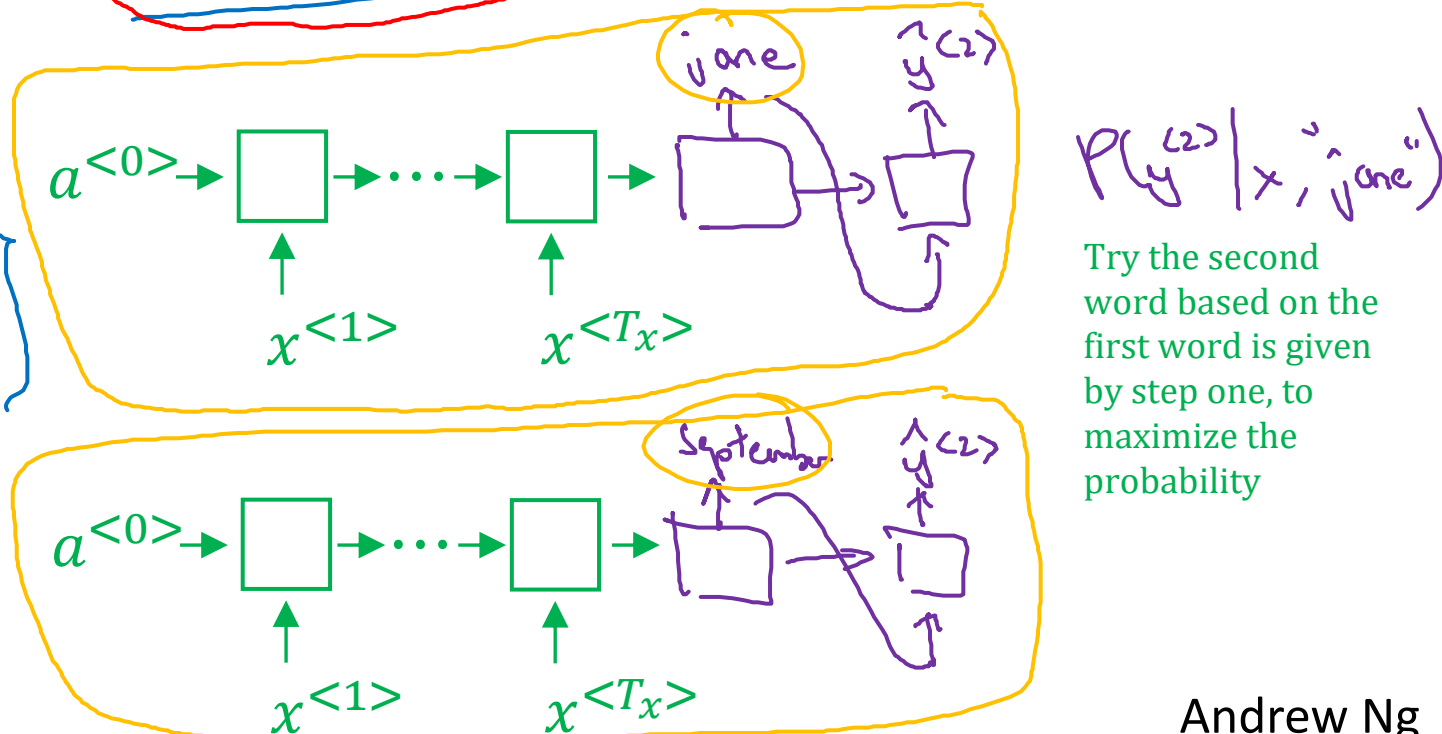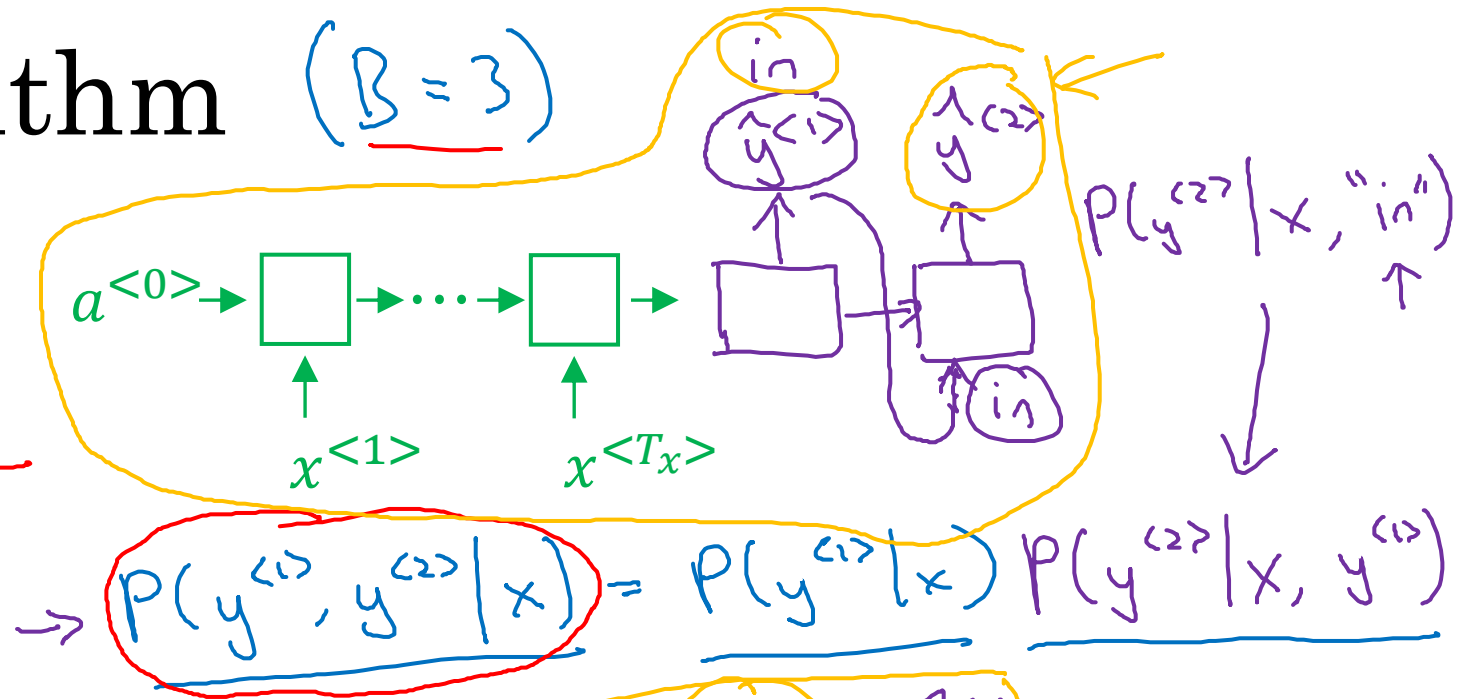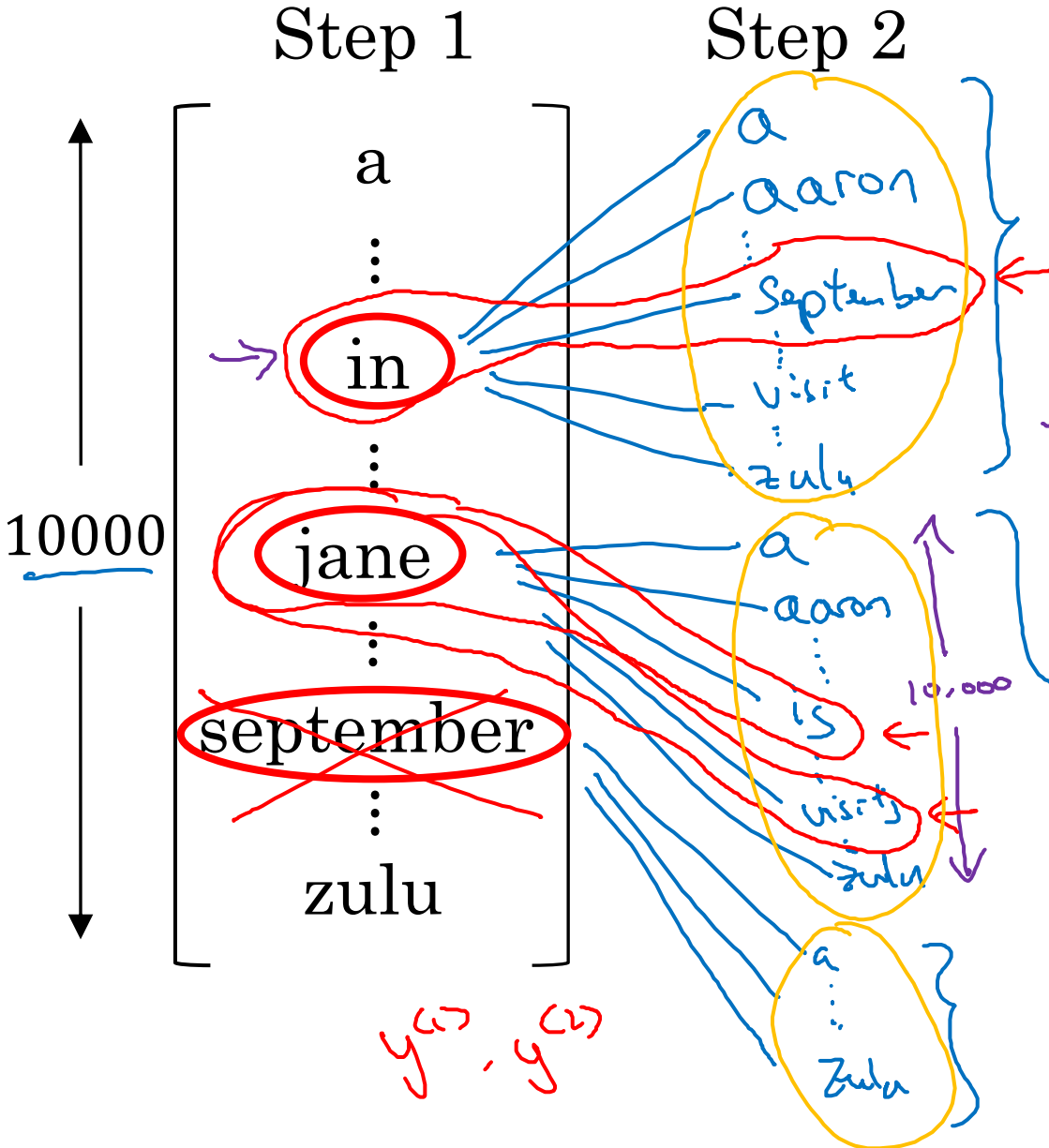
Beam search ($B = 3$)

$B = 1 \leadsto$ greedy search

in september
- a
- aaron
- jane
- zulu
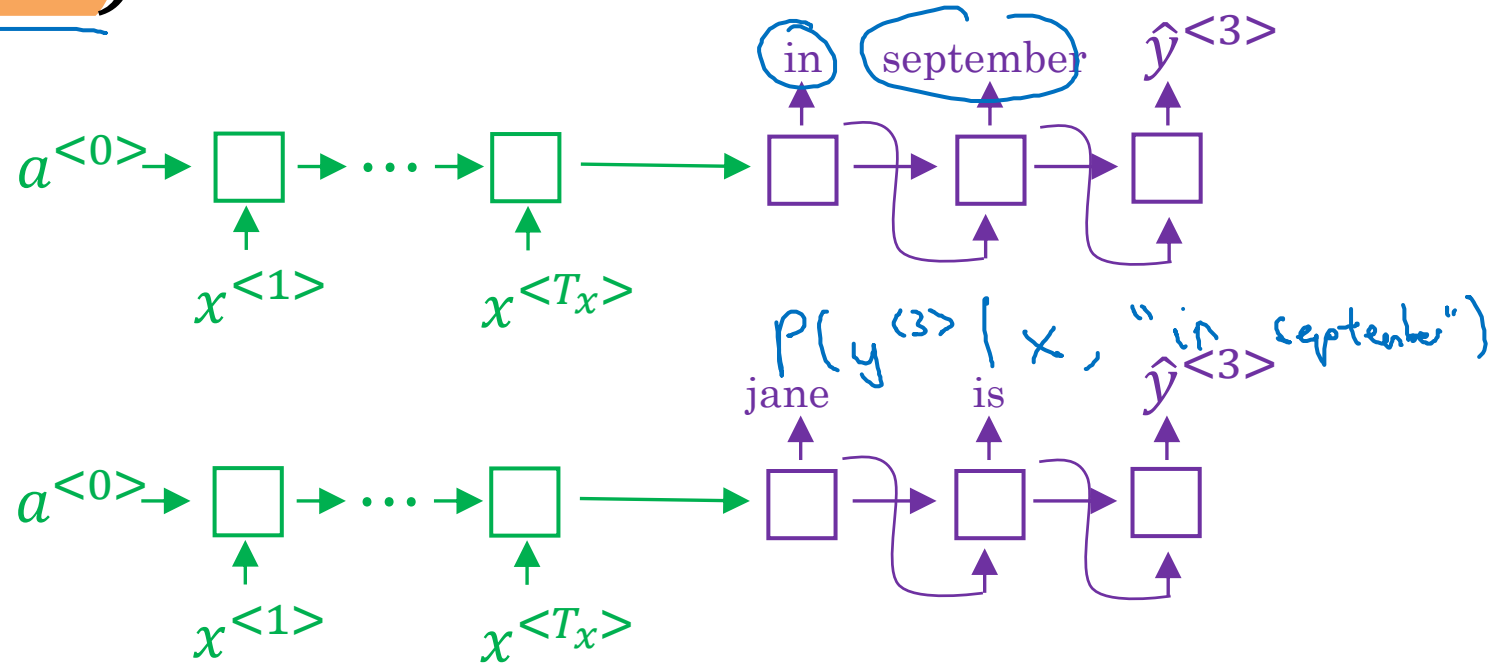
jane is
- a
- visits
- zulu

jane visits
- a
- africa
- zulu

in   september $\hat{y}^{<3>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$

$P(y^{<3>} | x, \text{"in september"})$

jane   is $\hat{y}^{<3>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$

jane   visits $\hat{y}^{<3>}$

$a^{<0>} \rightarrow \square \rightarrow \cdots \rightarrow \square \rightarrow \square \rightarrow \square \rightarrow \square$

$x^{<1>}$ $x^{<T_x>}$
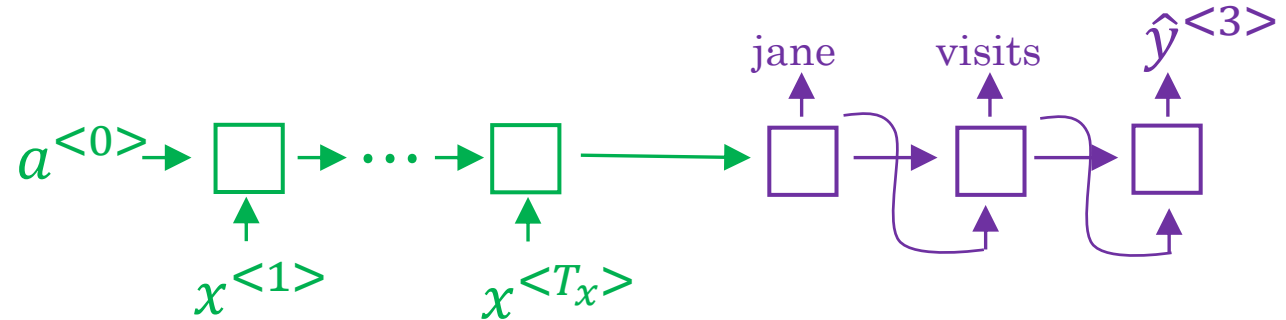
Try the third word, based on that the first two words are given, to maximize the probability.

$P(y^{<1>}, y^{<2>} | x)$

jane visits africa in september. <EOS>

Andrew Ng