كلية الهندسة
قسم هندسة الروبوت و الأنظمة الذكية

جامعة المنارة
MANARA UNIVERSITY

الشبكات العصبية

محاضرة 6

NLP and Word Embeddings

د. فادي متوج

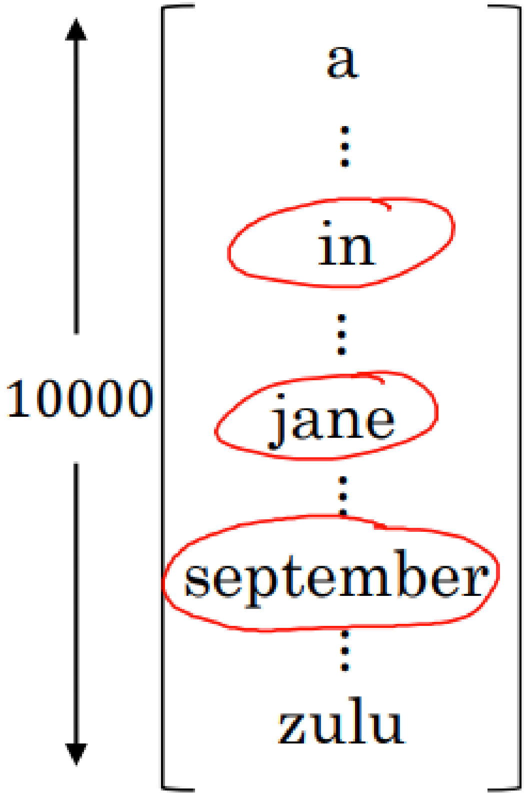
Sequence to sequence models

Beam search

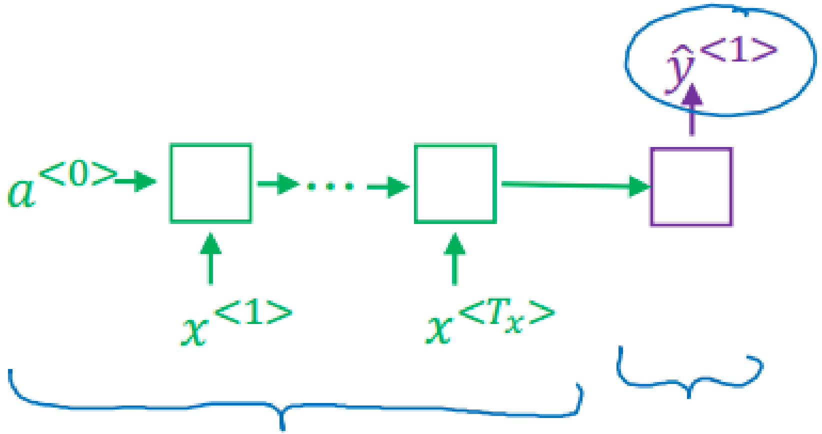
Beam search algorithm

$B = 3$ (beam width)

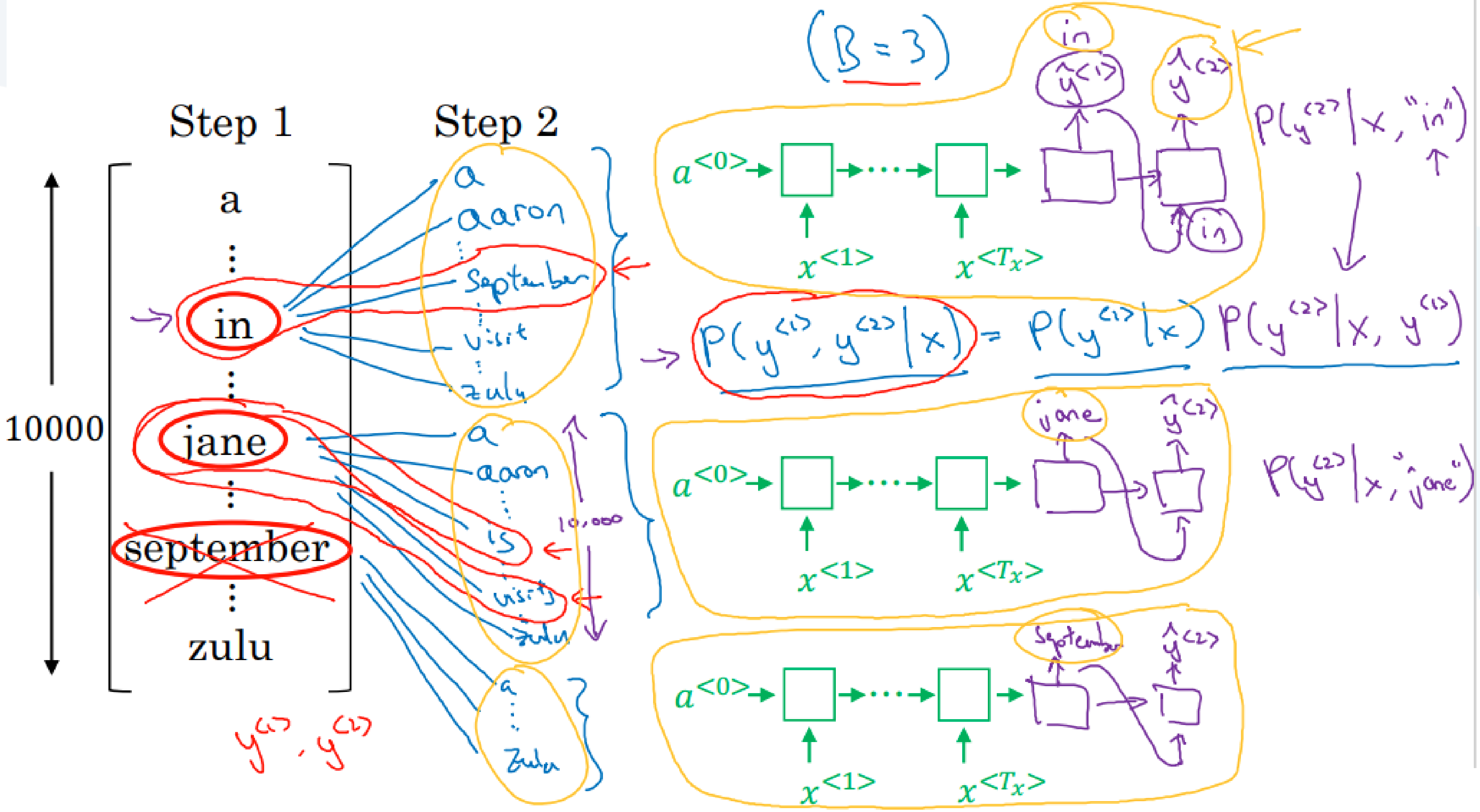
Step 1



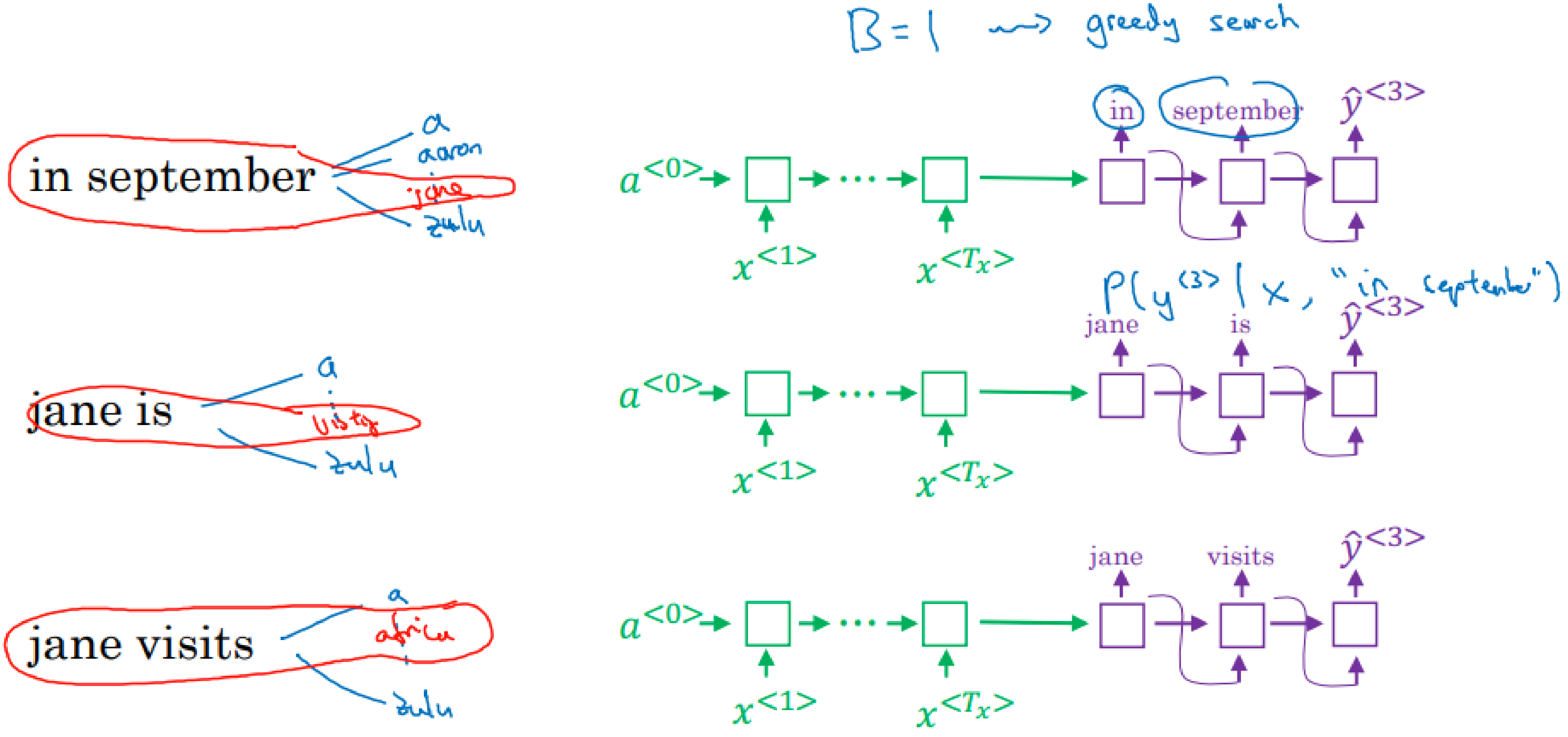
$\rightarrow P(y^{<1>} \mid x)$



Beam search algorithm



Beam search (4 = 3)



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

jane visits africa in september. <EOS>

Sequence to sequence models

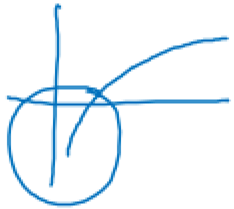
Refinements to Beam search

Length normalization

$P(y^{<1>} \dots y^{<T_y>} \mid x) = P(y^{<1>} \mid x)\, P(y^{<2>} \mid x, y^{<1>}) \dots P(y^{<T_y>} \mid x, y^{<1>}, \dots, y^{<T_y-1>})$

$$\arg\max_{y} \prod_{t=1}^{T_y} P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})$$

log



log $P(y|x)$ ←

$P(y|x)$ ←

$$\arg\max_{y} \sum_{t=1}^{T_y} \log P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>}) \leftarrow$$

$T_y = 1, 2, 3, \dots, 30.$

$$\rightarrow \boxed{\frac{1}{T_y^{\alpha}} \sum_{t=1}^{T_y} \log P(y^{<t>} \mid x, y^{<1>}, \dots, y^{<t-1>})}$$

$\alpha = 0.7$

$\alpha = 1$

$\alpha = 0$

Beam search discussion

large B: better result, slower
small B: worse result, faster

Beam width B?

1 → 3 → 10, 100, 1000 → 3000

Unlike exact search algorithms like BFS (Breadth First Search) or DFS (Depth First Search), Beam Search runs faster but is not guaranteed to find exact maximum for $\arg\max P(y/x)$

Sequence to sequence models

Error analysis on beam search

Example

→ RNN

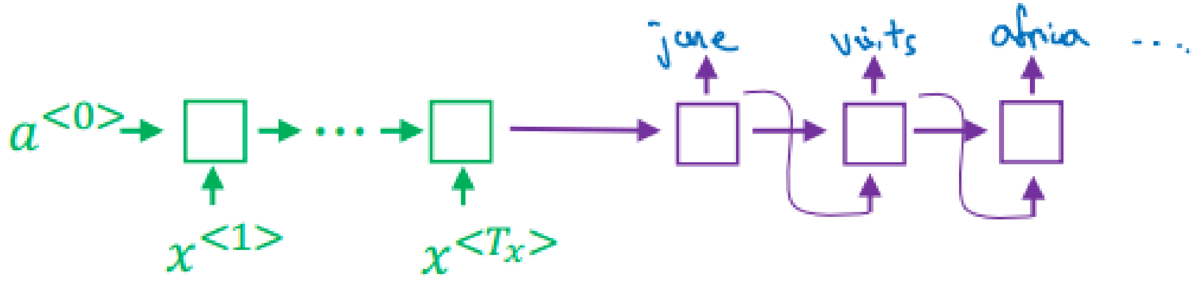
→ Beam Search

Jane visite l'Afrique en septembre.

Human: Jane visits Africa in September. (y^*)

Algorithm: Jane visited Africa last September. $(\hat{y})$ ←

RNN computes $P(y^*|x) \gtrless P(\hat{y}|x)$



Error analysis on beam search

$P(y^*|x)$

$P(\hat{y}|x)$

Human: Jane visits Africa in September. (y^*)

Algorithm: Jane visited Africa last September. ($\hat{y}$)

$\arg\max_y P(y|x)$

Case 1: $P(y^*|x) > P(\hat{y}|x) \leftarrow$

Beam search chose $\hat{y}$. But y^* attains higher $\boxed{P(y|x)}$.

Conclusion: Beam search is at fault.

Case 2: $P(y^*|x) \leq P(\hat{y}|x) \leftarrow$

y^* is a better translation than $\hat{y}$. But RNN predicted $\boxed{P(y^*|x)} < P(\hat{y}|x)$.

Conclusion: RNN model is at fault.

Error analysis process

Human	Algorithm	$P(y^*\|x)$	$P(\hat{y}\|x)$	At fault?
Jane visits Africa in September.	Jane visited Africa last September.	2×10^{-10}	1×10^{-10}	B
- - -	- - -	—	—	R
. . .	. . .	—	—	B
				R
				R
				⋮

Figures out what faction of errors are "due to" beam search vs. RNN model

Sequence to sequence models

BLUE (Bilingual Evaluation Understudy)

Evaluating machine translation

French: Le chat est sur le tapis.

Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: the the the the the the the.

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Bleu score on bigrams

French: Le chat est sur le tapis.

Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: the the the the the the the.

	Count	$Count_{clip}$
the cat	2	1
cat the	1	0
cat on	1	1
on the	1	1
the mat	1	1

$$\frac{4}{6}$$

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Bleu score on unigrams

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: The cat the cat on the mat.

$$P_1 \cdot P_2 \cdot = 1.0$$

$$(\hat{y})$$

$$P_1 = \frac{\sum_{unigrams \in \hat{y}} Count_{clip}(unigram)}{\sum_{unigram \in \hat{y}} Count(unigram)}$$

unigram

n-gram

$$P_n = \frac{\sum_{n\text{-}grams \in \hat{y}} Count_{clip}(n\text{-}gram)}{\sum_{n\text{-}gram \in \hat{y}} Count(n\text{-}gram)}$$

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Bleu details

p_n = Bleu score on n-grams only

p_1, p_2, p_3, p_4

Combined Bleu score $= BP \cdot \exp\left(\frac{1}{4}\sum_{n=1}^{4} p_n\right)$

BP: Brevity Penalty

$$BP = \begin{cases} 1 & \text{if MT\_output\_length} > \text{reference\_output\_length} \\ \exp(1 - \text{MT\_output\_length}/\text{reference\_output\_length}) & \text{otherwise} \end{cases}$$

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]