



deeplearning.ai

NLP and Word Embeddings

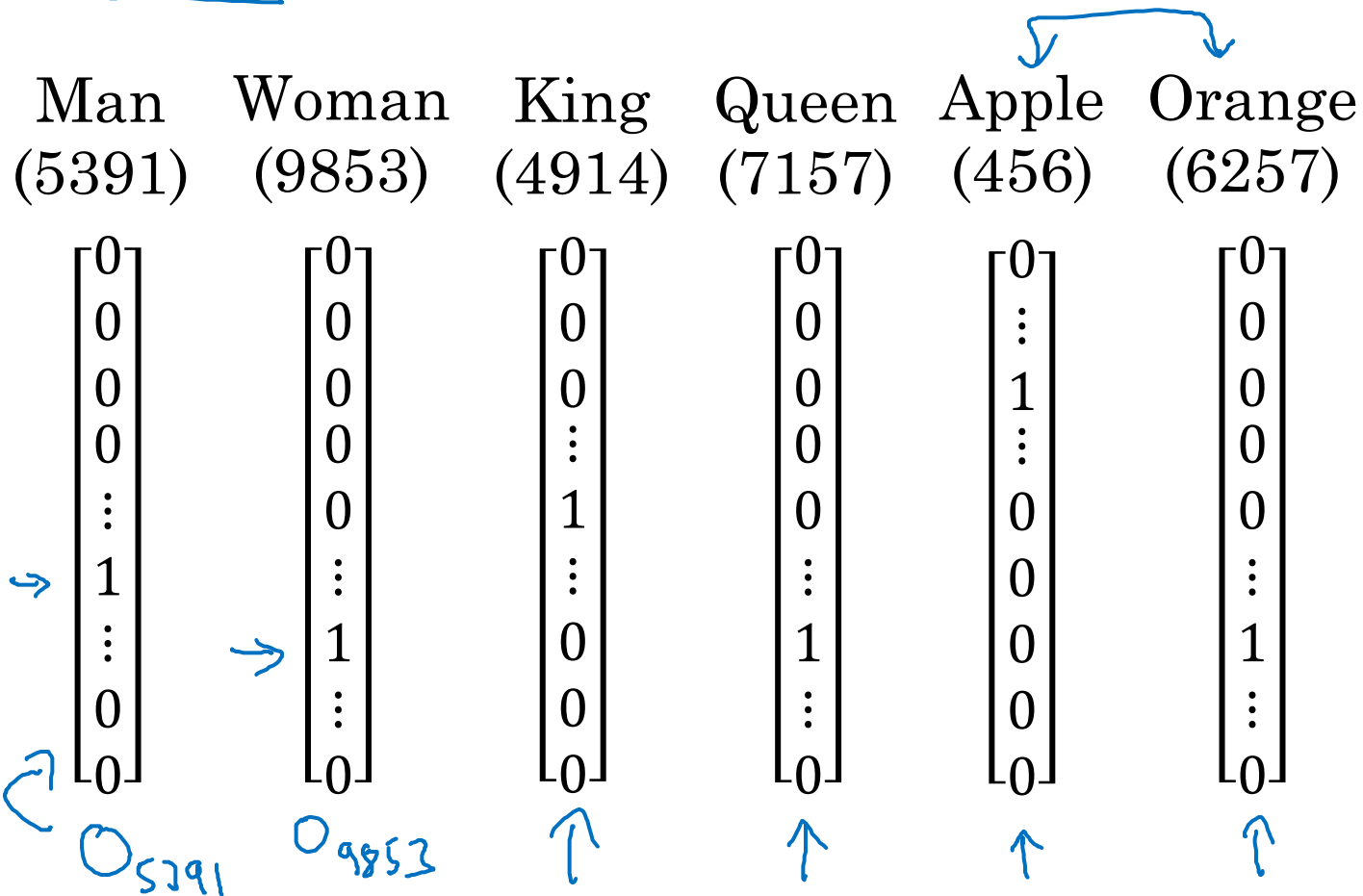
Word representation

Word representation

$V = [a, aaron, \dots, zulu, <UNK>]$

$|V| = 10,000$

1-hot representation



I want a glass of orange juice.

I want a glass of apple ?.

Featurized representation: word embedding

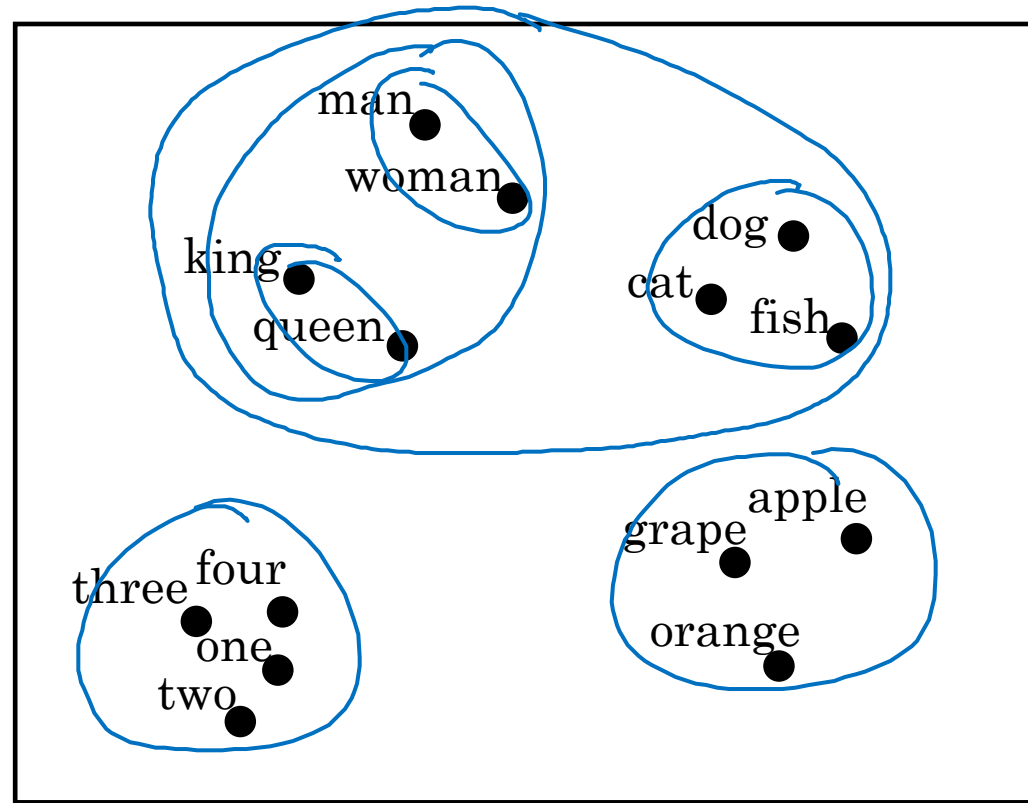
	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	<u>0.93</u>	<u>0.95</u>	-0.01	0.00
Age	0.03	0.02	0.7	0.69	0.03	-0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97
...				
size						
cost						
alive						
verb						

I want a glass of orange juice.

I want a glass of apple juice.

Andrew Ng

Visualizing word embeddings

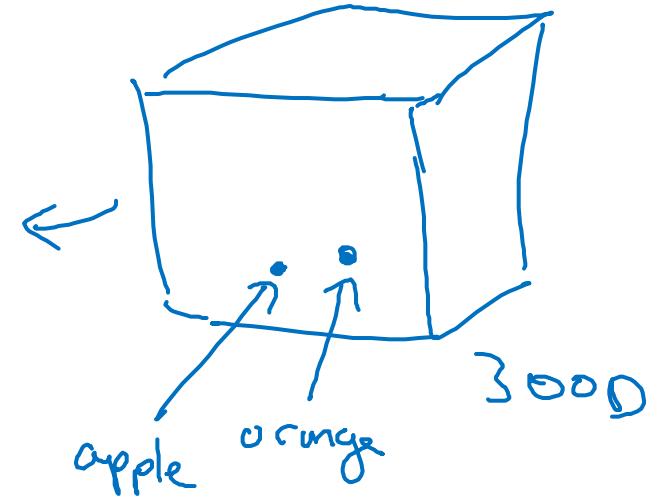


t-SNE

→ 300D



2D



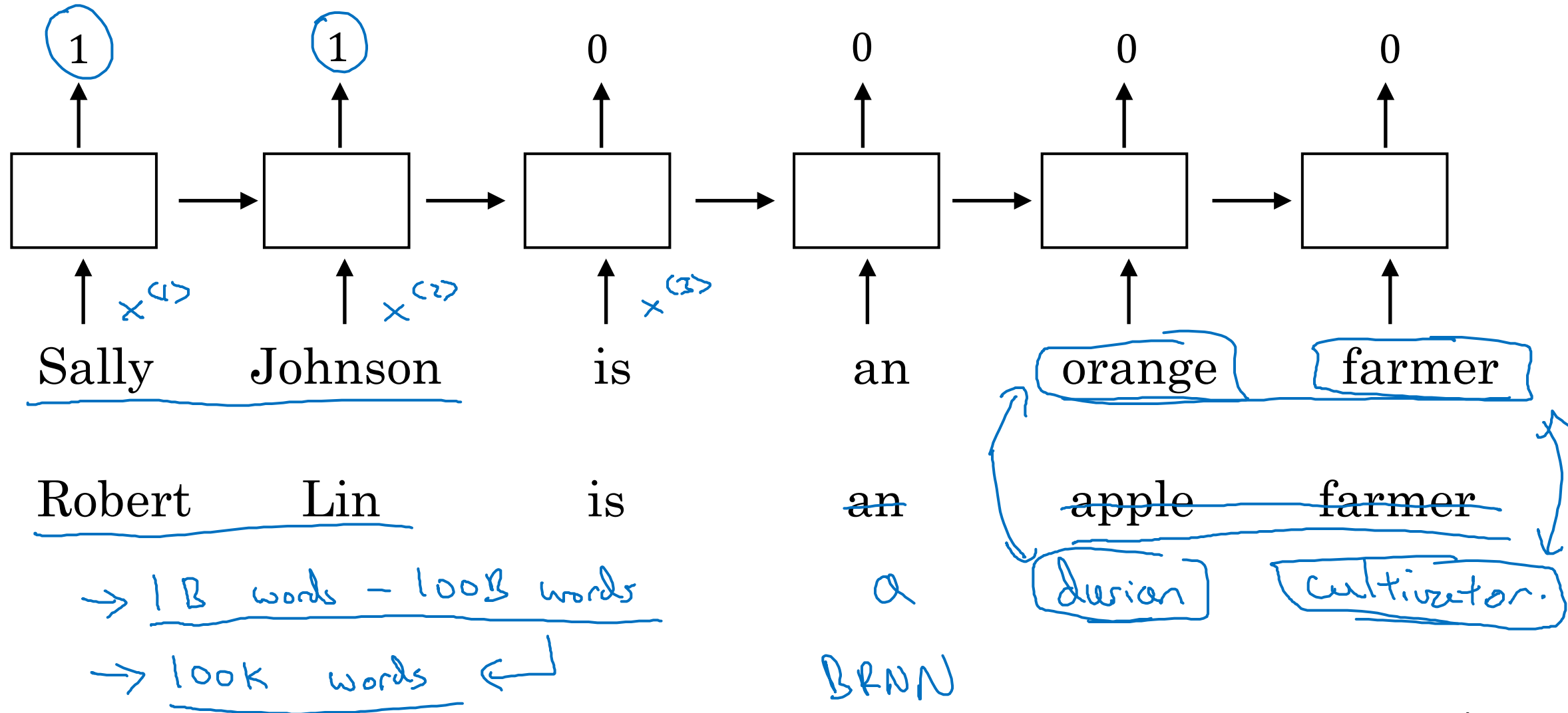


deeplearning.ai


NLP and Word Embeddings

Using word embeddings

Named entity recognition example



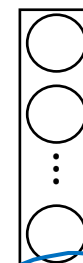
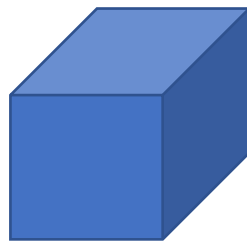
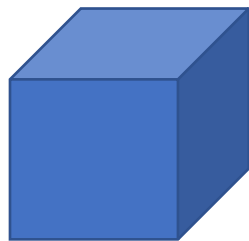
Transfer learning and word embeddings

- 
1. Learn word embeddings from large text corpus. (1-100B words)
(Or download pre-trained embedding online.)
 2. Transfer embedding to new task with smaller training set.
(say, 100k words) → 10,000 → 300
 3. Optional: Continue to finetune the word embeddings with new data.

Relation to face encoding (embedding) 128D



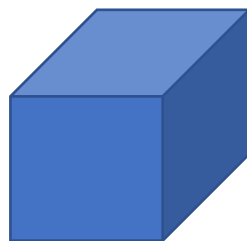
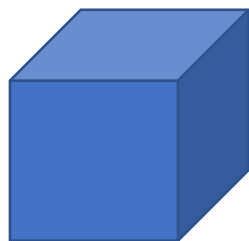
$x^{(i)}$



$f(x^{(i)})$



$x^{(j)}$



$f(x^{(j)})$



\hat{y}

$|V| = 10,000$

$e_1, \dots, e_{10,000}$



deeplearning.ai

NLP and Word Embeddings

Properties of word embeddings

Analogy

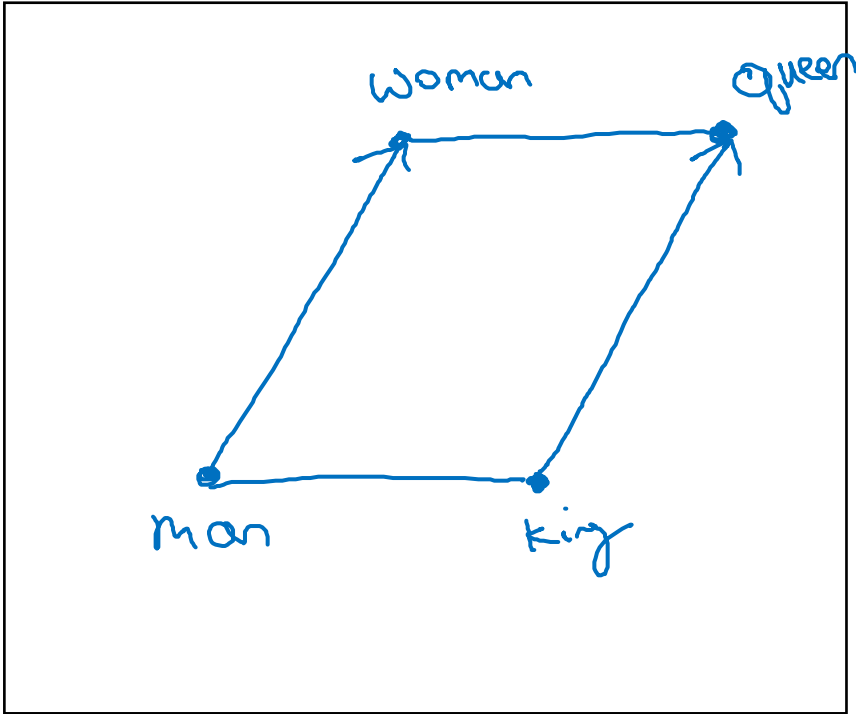
	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

$$\underbrace{e_{\text{man}}}_{5391} - \underbrace{e_{\text{woman}}}_{9853} \approx \underbrace{e_{\text{king}}}_{4914} - \underbrace{e_{\text{queen}}}_{7157}$$

$$\text{Man} \rightarrow \text{Woman} \quad \Leftrightarrow \quad \text{King} \rightarrow ? \text{ Queen}$$

$$\underbrace{e_{\text{man}} - e_{\text{woman}}}_{\approx \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}} \approx \underbrace{e_{\text{king}} - e_{\text{queen}}}_{\approx \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}}$$

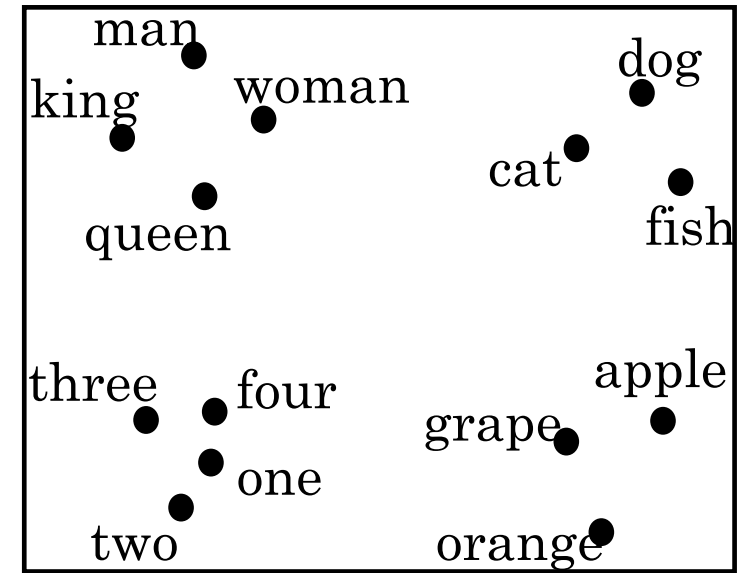
Analogies using word vectors



300 D

Find word w : $\arg \max_w$

3000 \rightarrow 20
↑



t-SNE

$$e_{man} - e_{woman} \approx e_{king} - \cancel{e_w} \quad \underline{e_w}$$

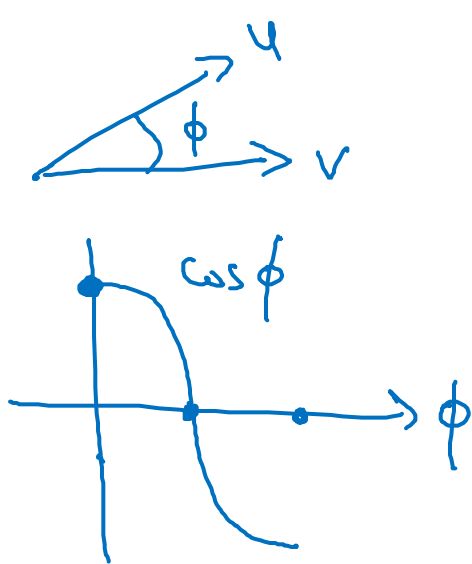
$$\text{Sim}(\underline{e_w}, \underline{e_{king} - e_{man} + e_{woman}})$$

30 - 75%

Cosine similarity

$$\rightarrow \text{sim}(e_w, e_{king} - e_{man} + e_{woman})$$

$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$$



$$\|u - v\|^2$$

Man:Woman as Boy:Girl

Ottawa:Canada as Nairobi:Kenya

Big:Bigger as Tall:Taller

Yen:Japan as Ruble:Russia

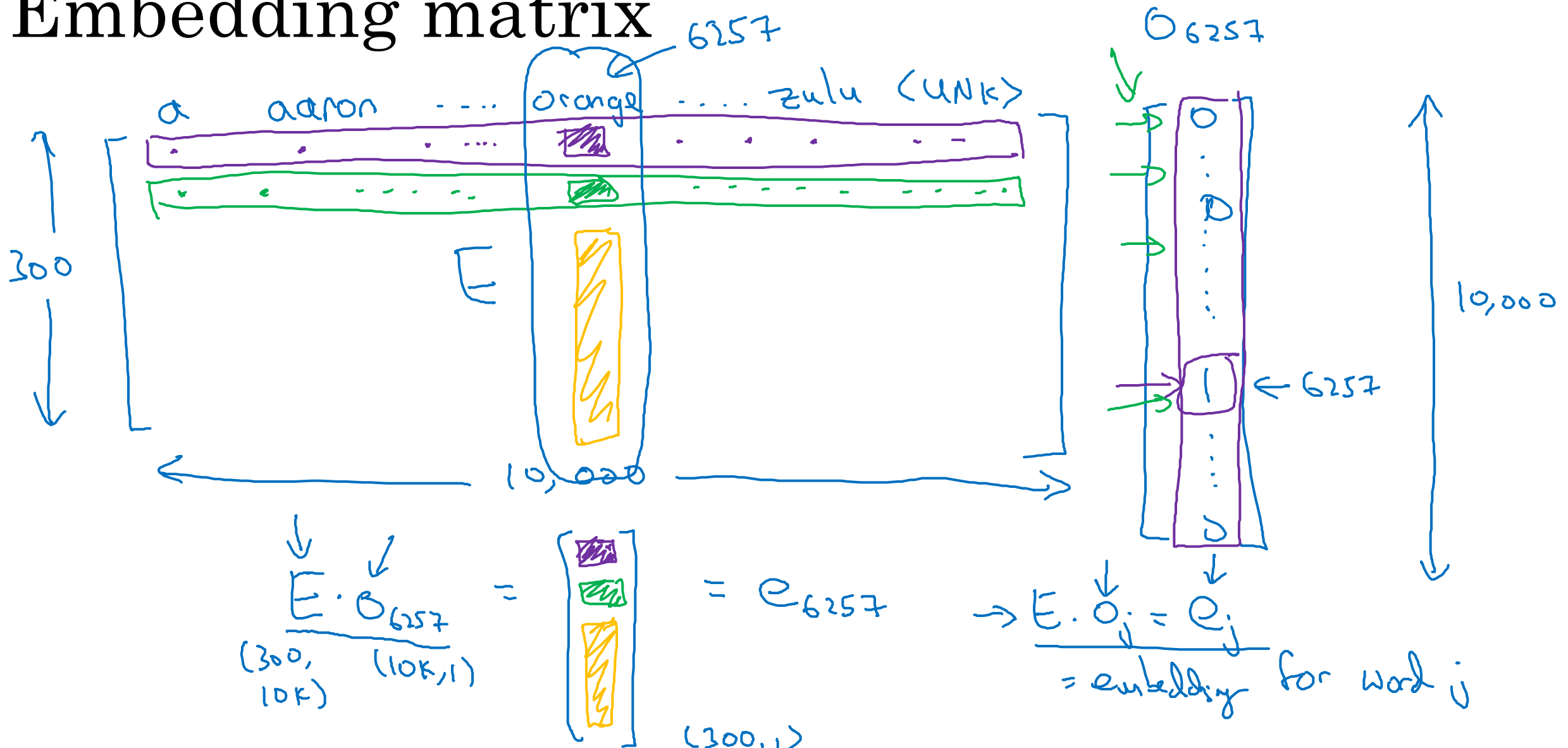


deeplearning.ai

NLP and Word Embeddings

Embedding matrix

Embedding matrix



In practice, use specialized function to look up an embedding.
 $\rightarrow \text{Embedding}$

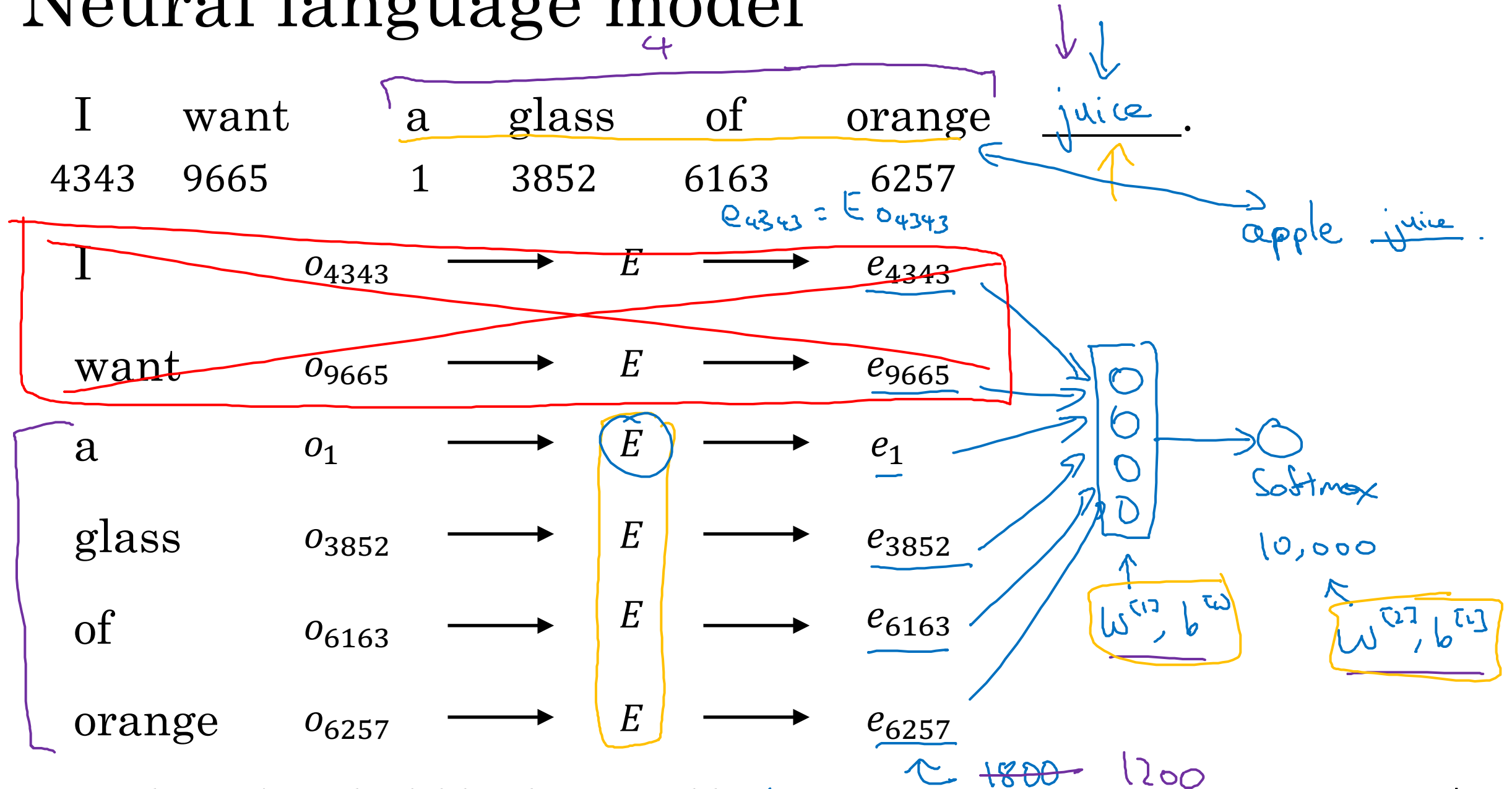


deeplearning.ai

NLP and Word Embeddings

Learning word embeddings

Neural language model



Other context/target pairs

I want a glass of orange juice to go along with my cereal.

The diagram illustrates the context and target for the word 'juice'. A purple bracket labeled 'context' spans the words 'a glass of orange'. A blue bracket labeled 'target' is positioned under the word 'juice'. A green arrow points from the word 'orange' to the word 'juice'.

Context: Last 4 words.

- 4 words on left & right
- Last 1 word
- Nearby 1 word

a glass of orange ? to go along with

orange ?

glass ?

skip gram



deeplearning.ai

NLP and Word Embeddings

Word2Vec

Skip-grams

I want a glass of orange juice to go along with my cereal.



Context

orange

orange

orange



Target

juice

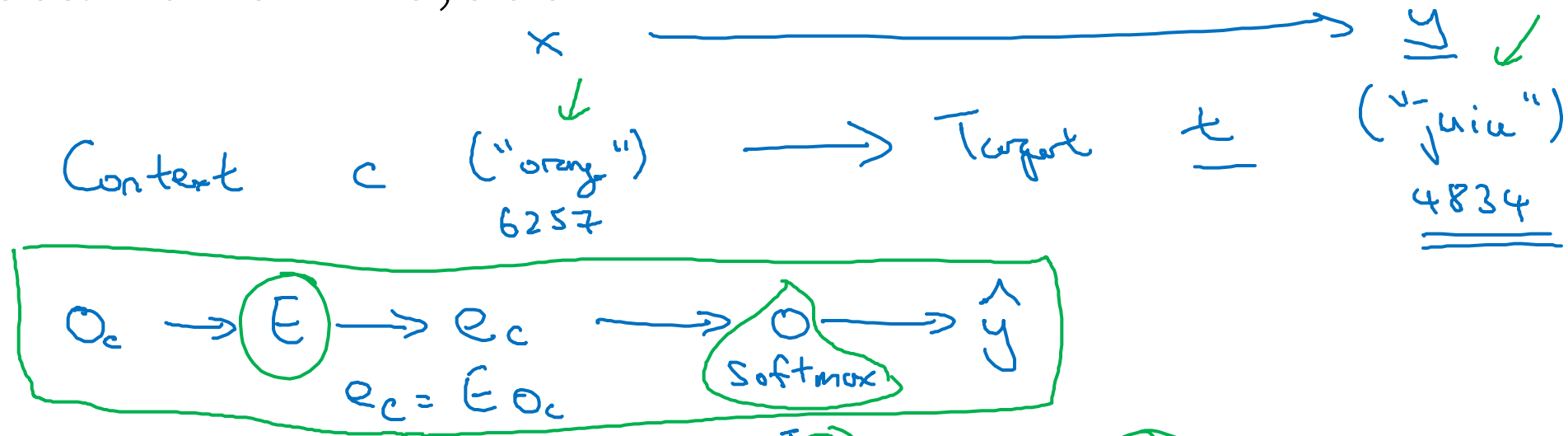
glass

my



Model

Vocab size = 10,000k



Softmax: $p(t|c) = \frac{e^{\theta_t^T e_c}}{\sum_{j=1}^{10,000} e^{\theta_j^T e_c}}$

θ_t = parameter associated with output t

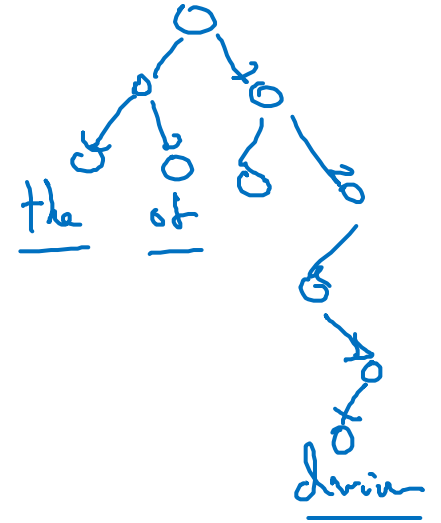
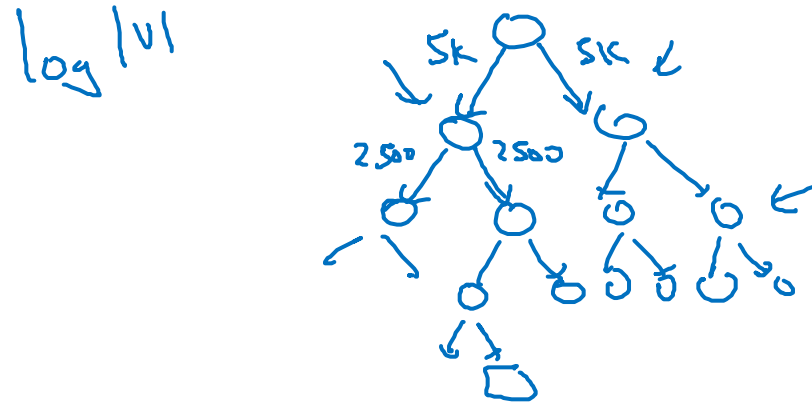
$\rightarrow \mathcal{L}(\hat{y}, y) = - \sum_{i=1}^{10,000} y_i \log \hat{y}_i$

$y = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow 4834$

Problems with softmax classification

$$\underline{p(t|c)} = \frac{e^{\theta_t^T \underline{e_c}}}{\sum_{j=1}^{10,000} e^{\theta_j^T e_c}}$$

Hierarchical softmax.



How to sample the context c ?

→ the, of, a, and, to, ...

→ orange, apple, divin

P_{divin}

t

$c \rightarrow t$

$P(c)$



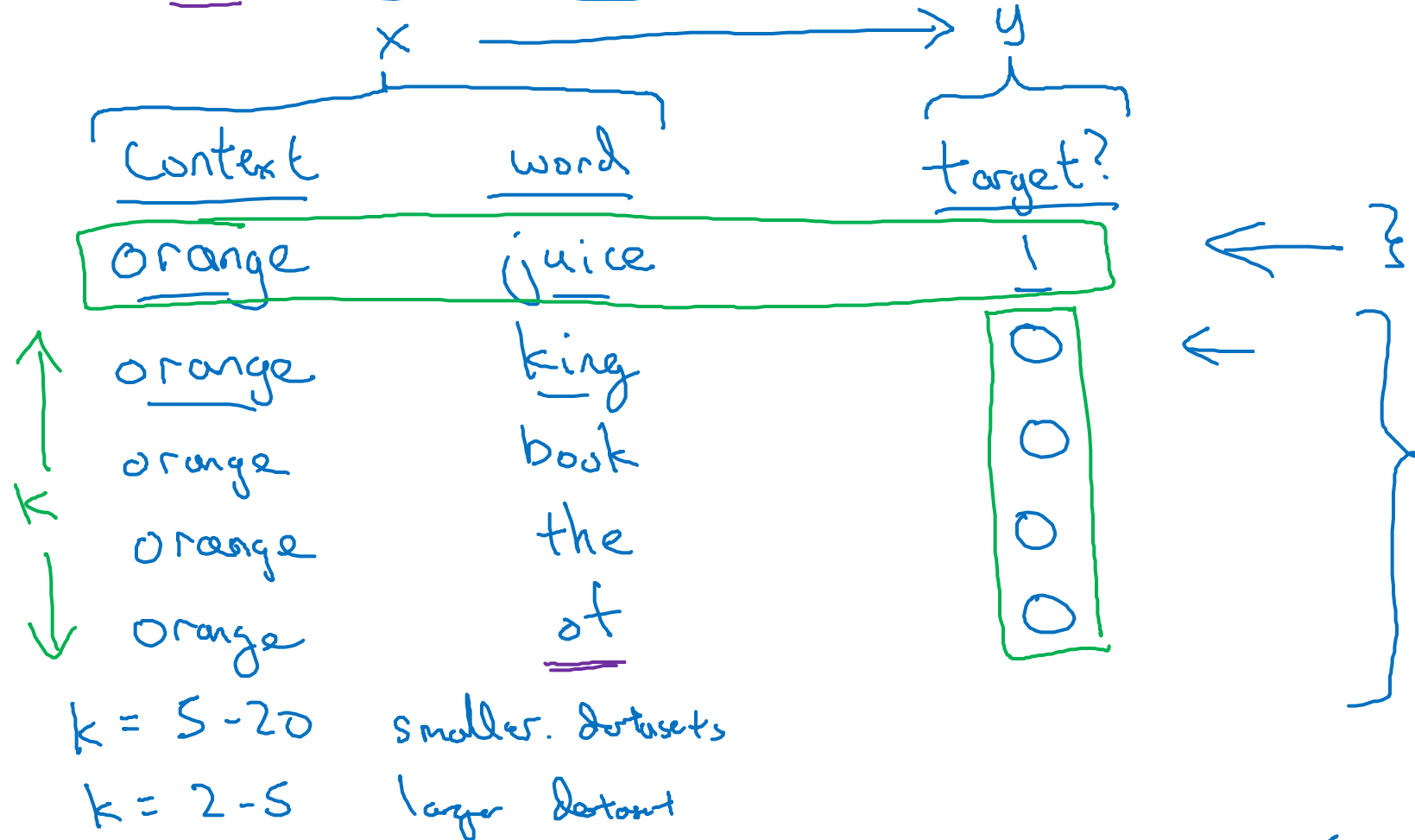
deeplearning.ai

NLP and Word Embeddings

Negative sampling

Defining a new learning problem

I want a glass of orange juice to go along with my cereal.



Model

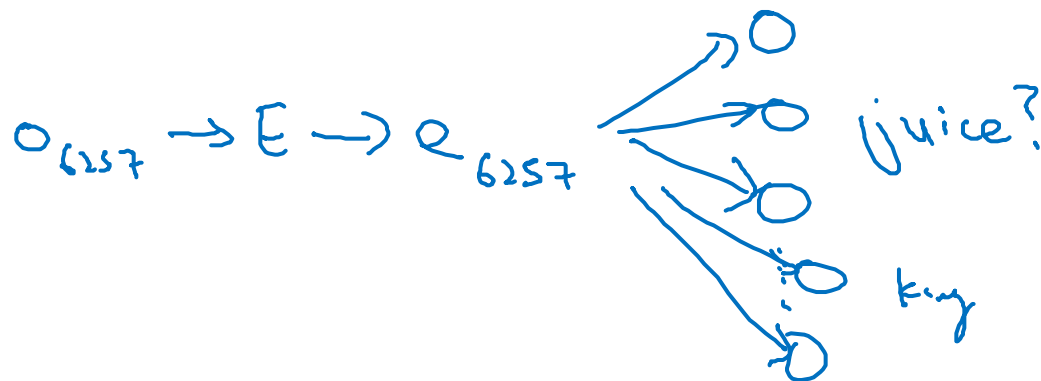
Softmax:
$$p(t|c) = \frac{e^{\theta_t^T e_c}}{\sum_{j=1}^{10,000} e^{\theta_j^T e_c}}$$

10,000-way softmax

$$P(y=1 | c, t) = \sigma(\theta_t^T e_c) \leftarrow$$

x		y
<u>context</u>	<u>word</u>	<u>target?</u>
orange	juice	1
orange	king	0
orange	book	0
orange	the	0
orange	of	0
\uparrow c	\uparrow t	\uparrow y

Orange
6257



\uparrow
10,000
 \downarrow

10,000 binary classification problem

$k+1$

Selecting negative examples

<u>context</u>	<u>word</u>	<u>target?</u>
orange	juice	1
orange	king	0
orange	book	0
orange	the	0
orange	of	0

↑
t

the, of, and, ...

$$P(w_i) = \frac{f(w_i)^{3/4}}{\sum_{j=1}^{10,000} f(w_j)^{3/4}}$$

$$\frac{1}{|V|}$$

↑



deeplearning.ai

NLP and Word Embeddings

GloVe word vectors

GloVe (global vectors for word representation)

I want a glass of orange juice to go along with my cereal.

c, t

X_{ij} = # times i appears in context of j .

$\begin{matrix} \uparrow & \uparrow \\ c & t \end{matrix}$ $\begin{matrix} \uparrow \\ t \end{matrix}$ $\begin{matrix} \uparrow \\ c \end{matrix}$

$X_{ij} = X_{ji} \leftarrow$

Model

minimize

$$\sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(x_{ij}) \left(\underbrace{\Theta_i^T e_j}_{\substack{t \quad c \\ \text{"}\Theta_t^T e_c\text{"}}} + b_i + b_j' - \log x_{ij} \right)^2$$

weighting term

Annotations: Green arrows point to $\Theta_i^T e_j$, b_i , and b_j' . A blue arrow points from the weighting term to the log term. A green arrow points to the entire expression.

$f(x_{ij}) = 0$ if $x_{ij} = 0$.

" $0 \log 0$ " = 0

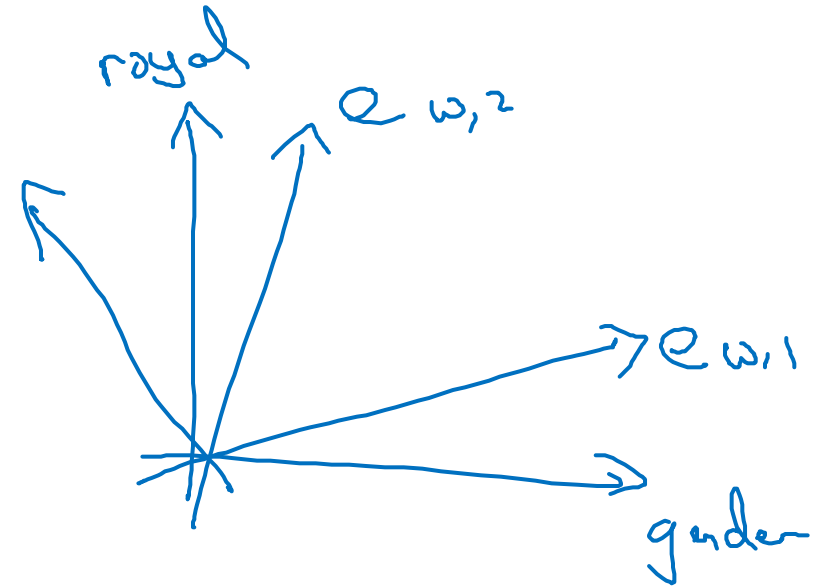
this, is, of, a, ...
derivation

Θ_i, e_j are symmetric

$$e_w^{(final)} = \frac{e_w + \Theta_w}{2}$$

A note on the featurization view of word embeddings

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	
Gender	-1	1	-0.95	0.97	←
Royal	0.01	0.02	0.93	0.95	←
Age	0.03	0.02	0.70	0.69	←
Food	0.09	0.01	0.02	0.01	←



$$\text{minimize } \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (\underbrace{\theta_i^T e_j}_{\text{handwritten}} + b_i - b'_j - \log X_{ij})^2$$

$$\underbrace{(A\theta_i)^T (A^{-T}e_j)}_{\text{handwritten}} = \theta_i^T \cancel{A^T A} e_j$$



deeplearning.ai

NLP and Word Embeddings

Sentiment classification

Sentiment classification problem



The dessert is excellent.



Service was quite slow.



Good for a quick meal, but nothing special.



Completely lacking in good taste, good service, and good ambience.



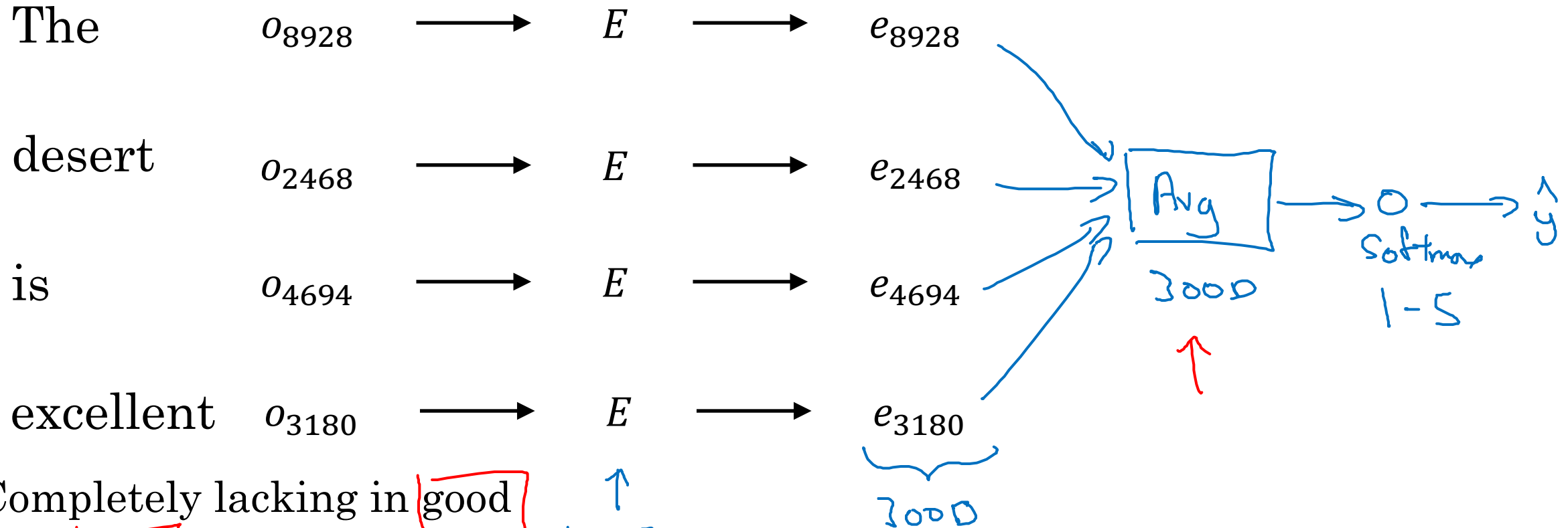
10,000  100,000 words

Simple sentiment classification model

The dessert is excellent



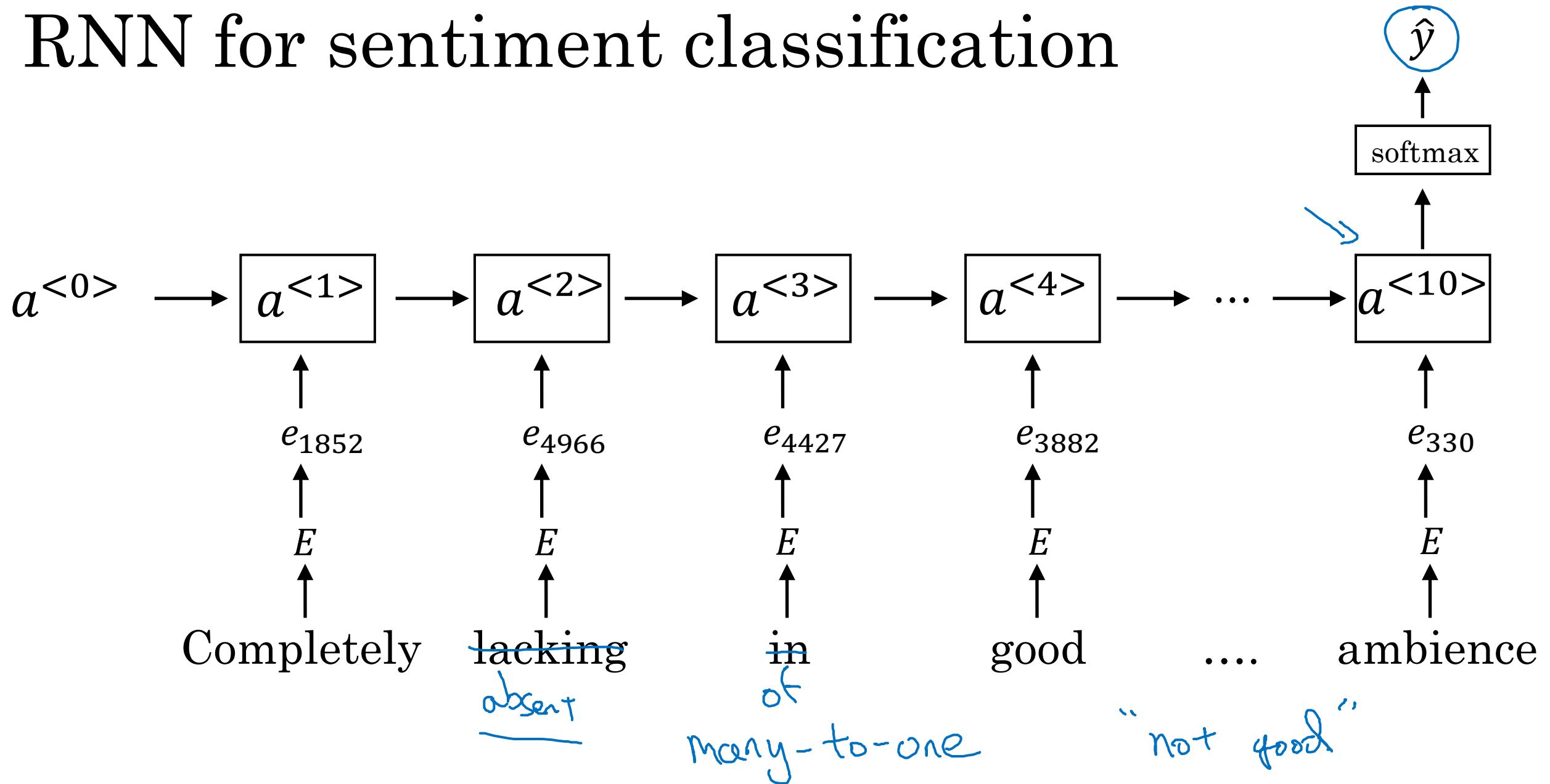
8928 2468 4694 3180



“Completely lacking in good taste, good service, and good ambience.”

↑
100 B
words

RNN for sentiment classification





deeplearning.ai

NLP and Word Embeddings

Debiasing word embeddings

The problem of bias in word embeddings

Man:Woman as King:Queen

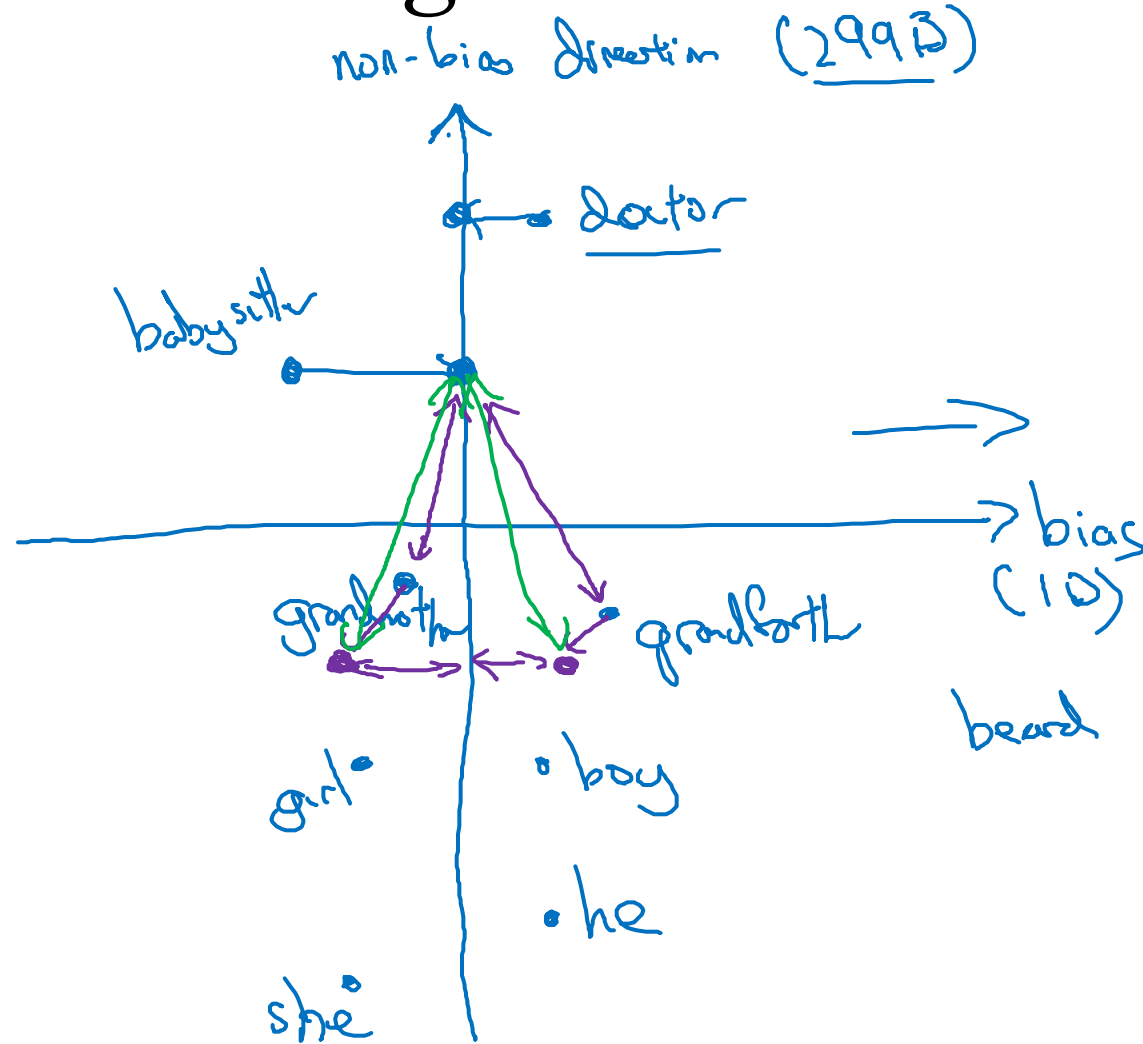
Man:Computer_Programmer as Woman:Homemaker X

Father:Doctor as Mother:Nurse X

Word embeddings can reflect gender, ethnicity, age, sexual orientation, and other biases of the text used to train the model.



Addressing bias in word embeddings



1. Identify bias direction.

$$\begin{cases} e_{he} - e_{she} \\ e_{male} - e_{female} \\ \vdots \end{cases} \rightarrow \text{average}$$

2. Neutralize: For every word that is not definitional, project to get rid of bias.

3. Equalize pairs.

$$\rightarrow \left. \begin{matrix} \text{grandmother} & - & \text{grandfather} \\ \text{girl} & & \text{boy} \end{matrix} \right\}$$