



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

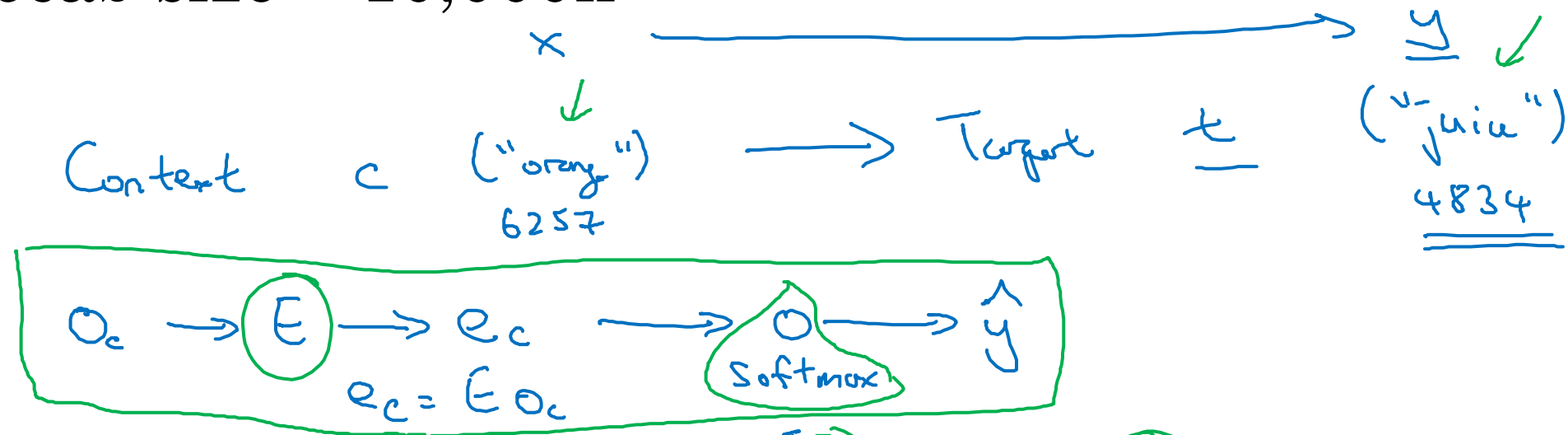
glass

my



# Model

Vocab size = 10,000k



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

Where  $\theta_t$  is the parameter associated with output  $t$ .

Loss function:

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

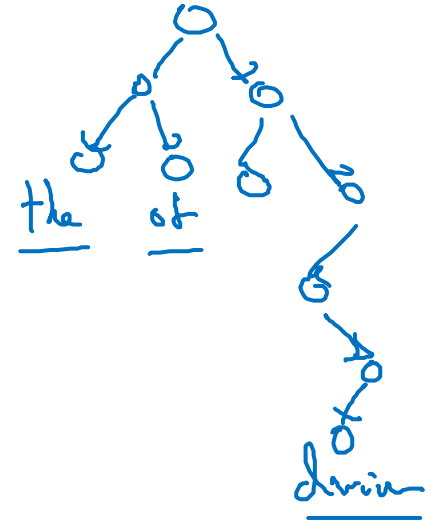
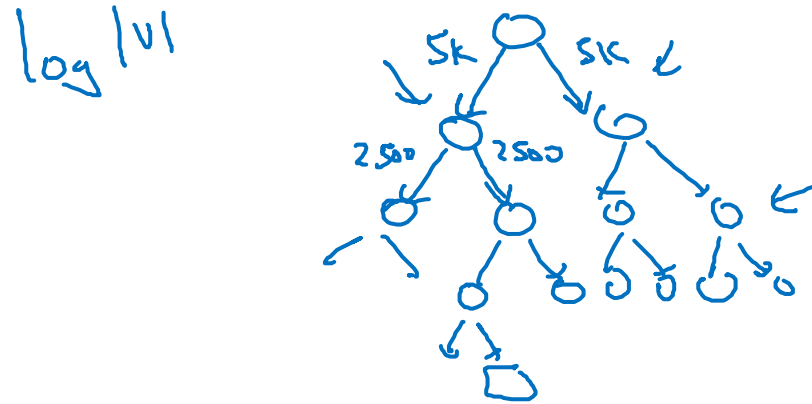
Output vector  $y$ :

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

# 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

$P_{divin}$

$t$

$c \rightarrow t$

$P(c)$