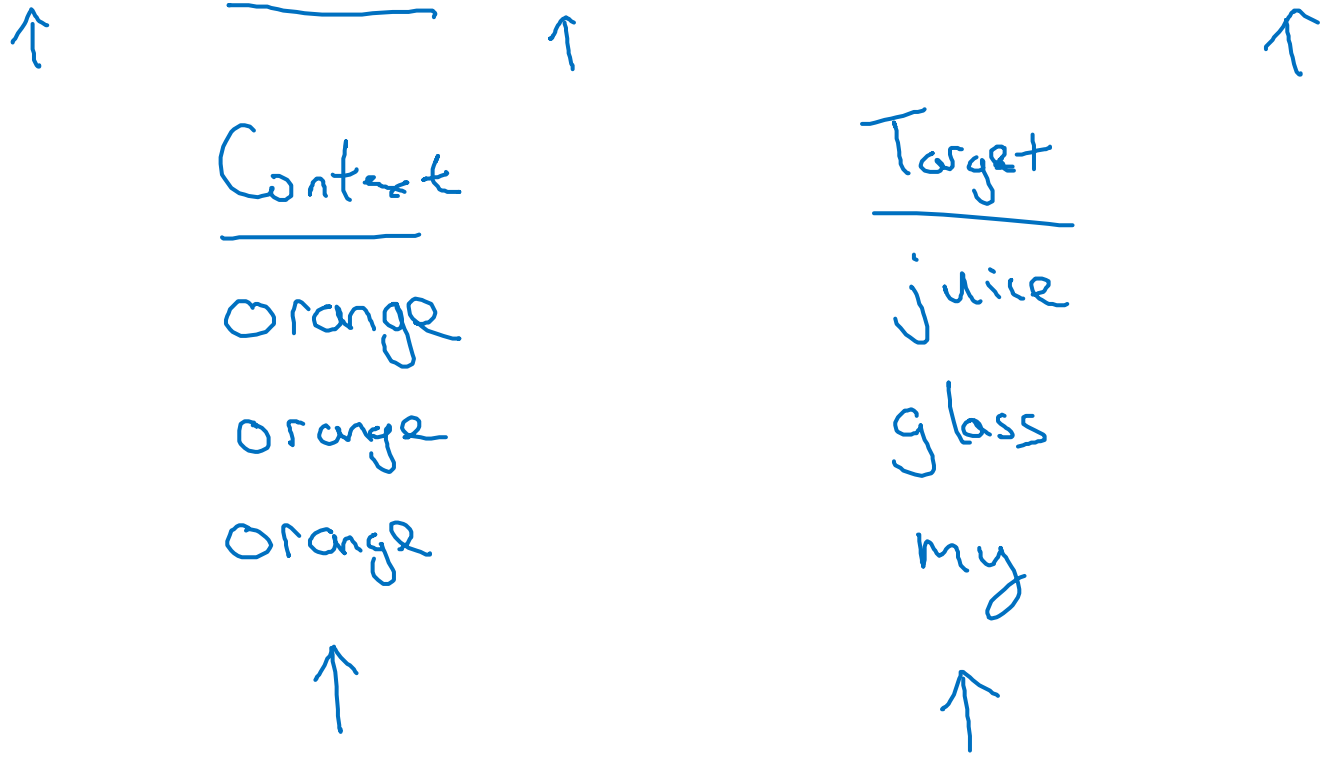
deeplearning.ai

NLP and Word Embeddings

Word2Vec

Skip-grams

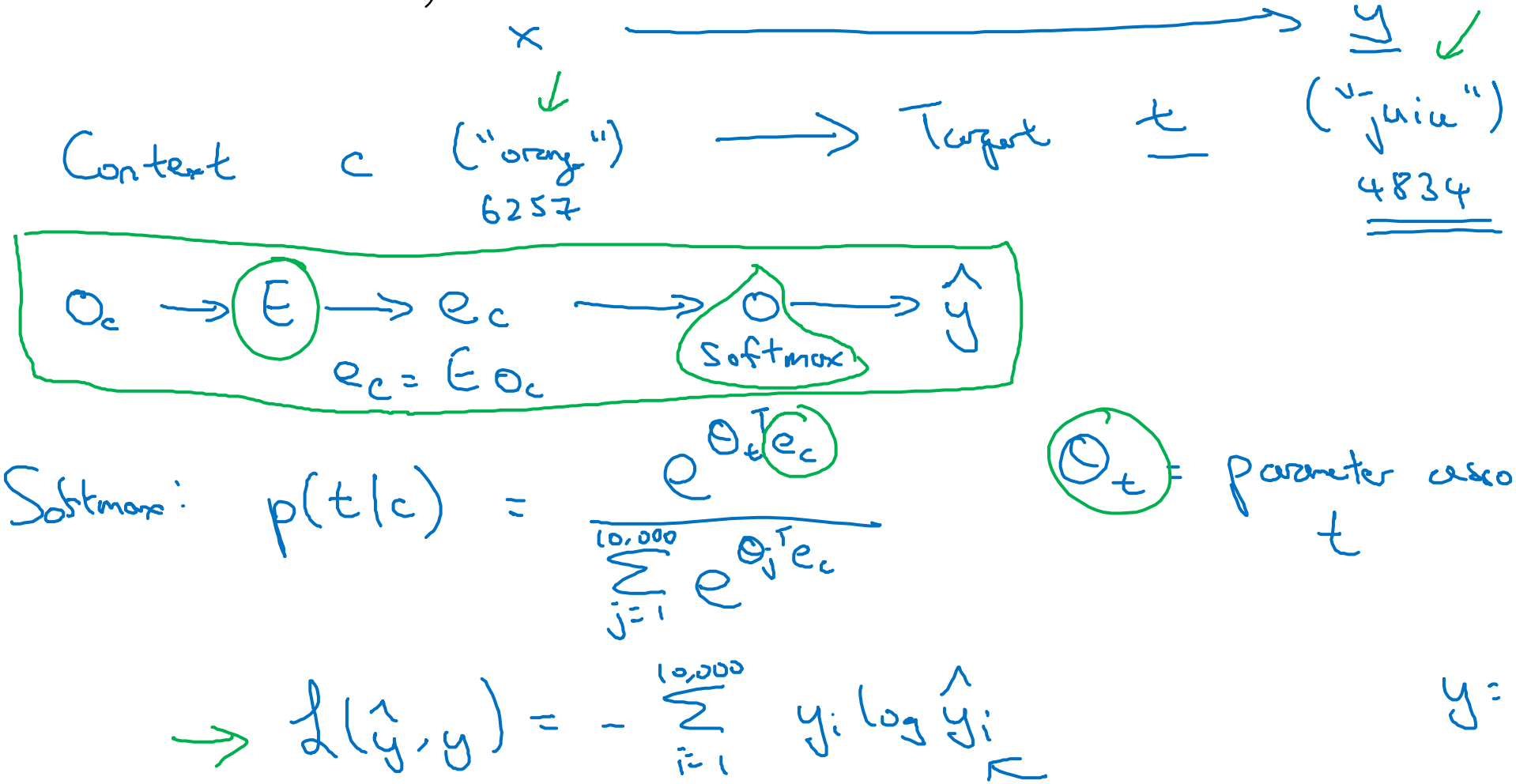
I want a glass of orange juice to go along with my cereal.



[Mikolov et. al., 2013. Efficient estimation of word representations in vector space.]

Model

Vocab size = 10,000k

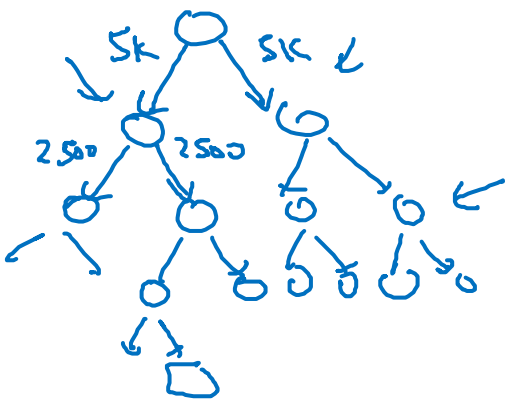


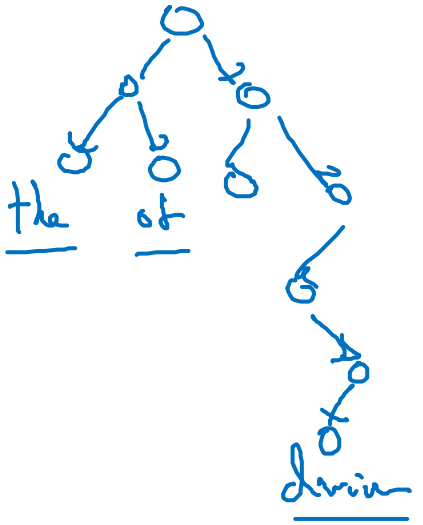
Problems with softmax classification

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

Hierarchical softmax.

$\log |V|$





How to sample the context c?

t

$c \to t$

→ the, of, a, and, to, ...

→ orange, apple, durian

$P(c)$

$\neq$ durian