NLP and Word Embeddings

Word2Vec

deeplearning.ai

Skip-grams

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

Context	Target
Orange	juice
Orange	glass
Orange	my

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

Andrew Ng

Model

Vocab size = 10,000k

x

Context c ("orange") $\longrightarrow$ Target t

6257

y ("juice")

4834

$$O_c \longrightarrow \boxed{E} \longrightarrow e_c \longrightarrow \text{Softmax} \longrightarrow \hat{y}$$

$e_c = E O_c$

$\theta_t^T e_c$

θ_t = parameter associated with output t
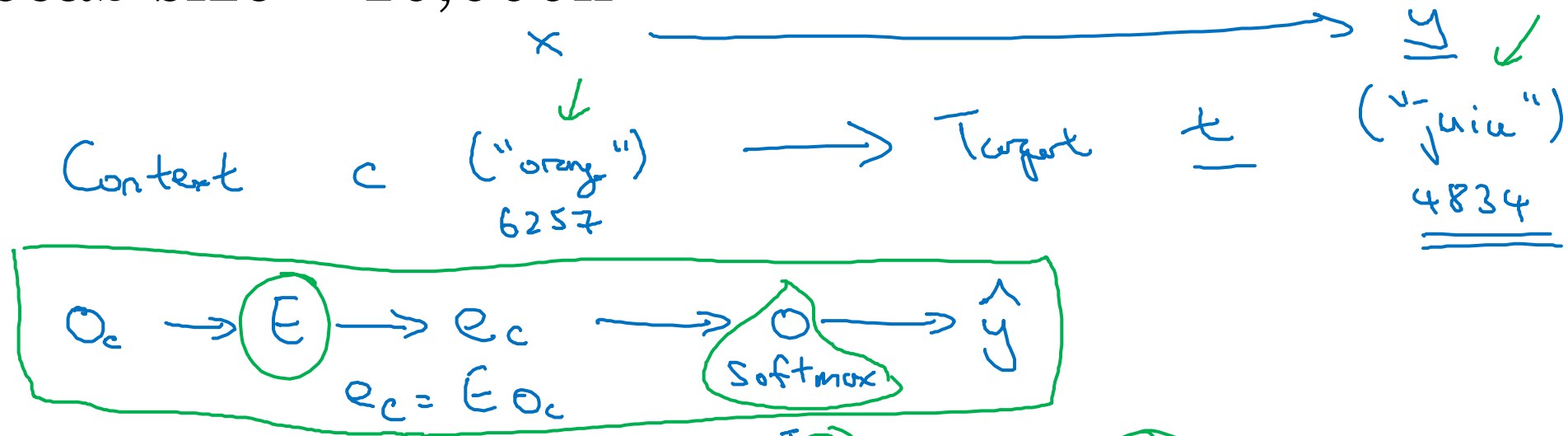
Softmax: $p(t|c) = \dfrac{e^{\theta_t^T e_c}}{\sum_{j=1}^{10,000} e^{\theta_j^T e_c}}$

$\rightarrow \mathcal{L}(\hat{y}, y) = -\sum_{i=1}^{10,000} y_i \log \hat{y}_i$
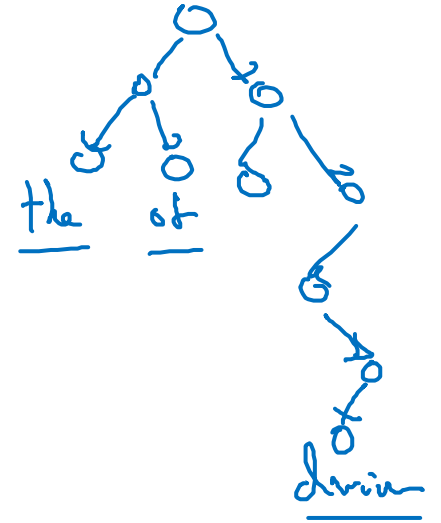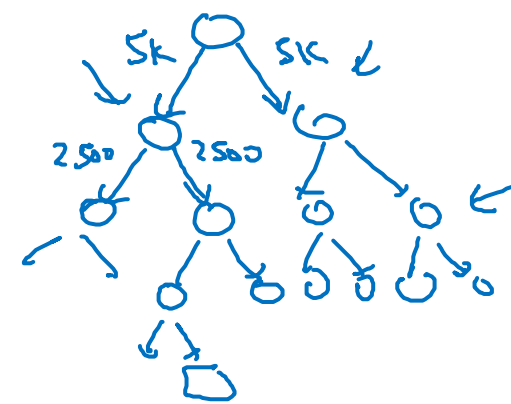
$$y = \begin{bmatrix} 0 \\ \vdots \\ i \\ \vdots \\ 0 \end{bmatrix} \leftarrow 4834$$

Andrew Ng

Problems with softmax classification

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

Hierachial   softmax.

$\log |v|$

5K   5K

2500   2500

the   of

durian

How to sample the context c?

$\rightarrow$ the, of, a, and, to, ...

$\rightarrow$ orange, apple, durian

2 durian

t

$c \rightarrow t$

$P(c)$