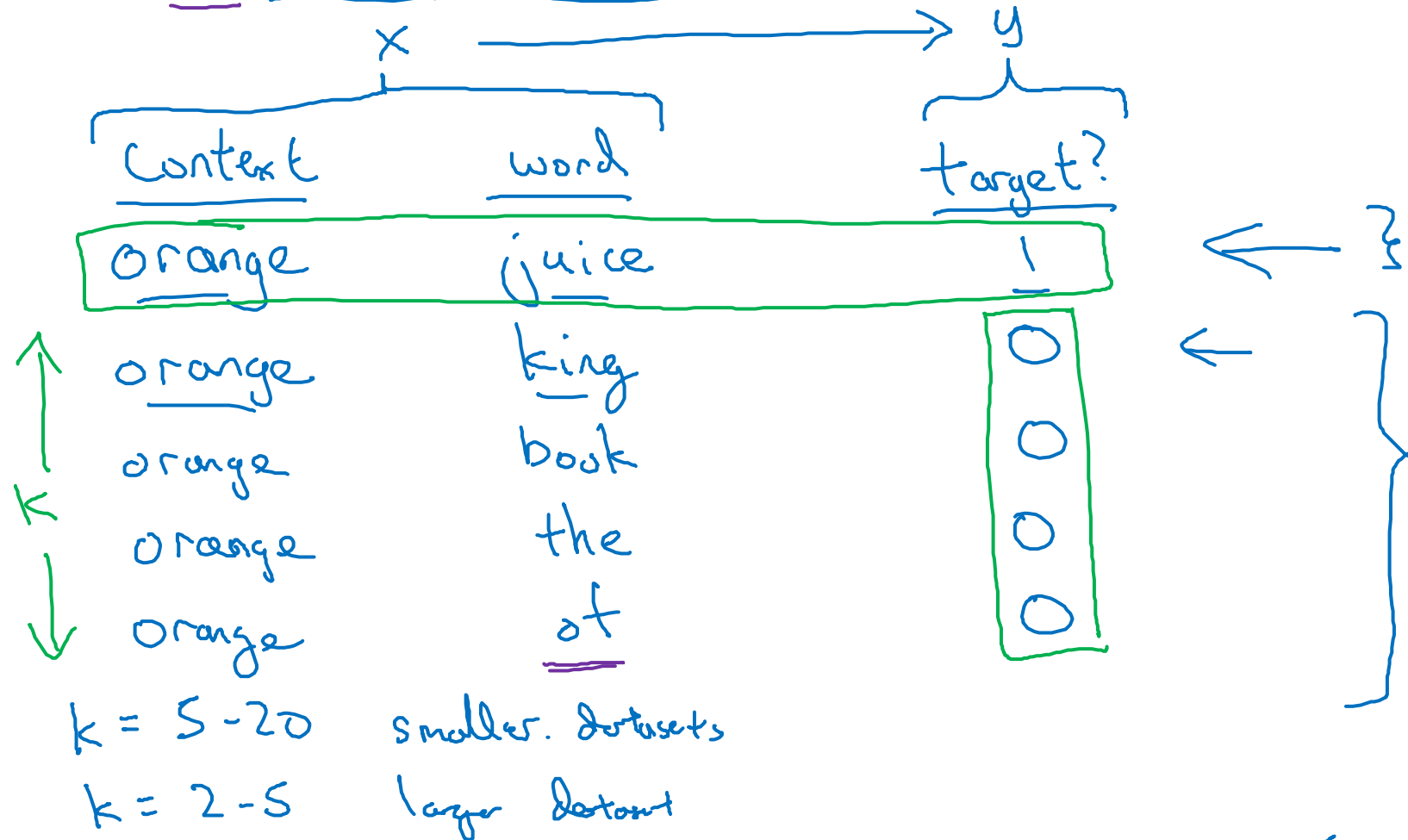NLP and Word Embeddings

Negative sampling

deeplearning.ai

Defining a new learning problem

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

Context	word	target?
orange	juice	1
orange	king	
orange	book	
orange	the	
orange	of	

$k = 5 - 20$ smaller datasets
$k = 2 - 5$ larger dataset

[Mikolov et. al., 2013. Distributed representation of words and phrases and their compositionality]
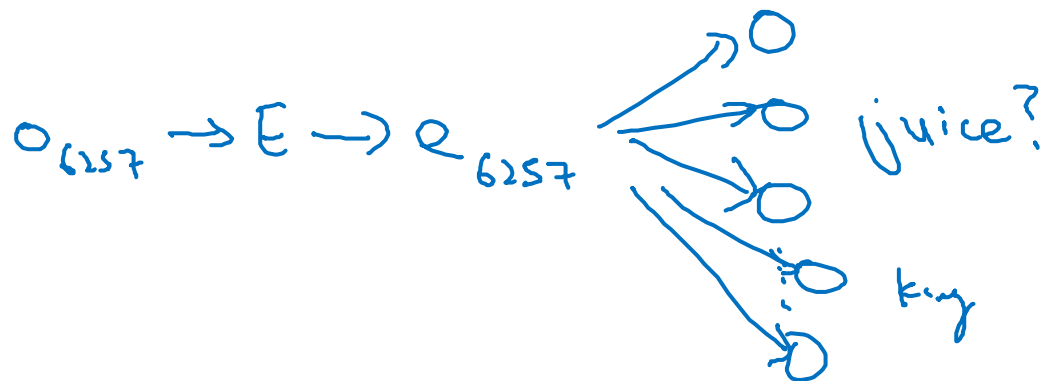
Andrew Ng

Model

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

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

Orange
6257

$O_{6257} \rightarrow E \rightarrow e_{6257}$

juice?

king

10,000

context	word	target?
orange	juice	1
orange	king	0
orange	book	0
orange	the	0
orange	of	0

x y

c t y

10,000 binary
classification
problem

$k+1$

Selecting negative examples

t

y

context	word	target?
orange	juice	1
orange	king	0
orange	book	0
orange	the	0
orange	of	0

t

the , of, and, …

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

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

Andrew Ng