Embedding matrix

6257

$\mathbf{O}_{6257}$ 网易云课堂

a   aaron   .....   Orange   .....   zulu   <UNK>

300

E

$(10,020)$

10,000

← 6257

$$E \cdot O_{6257} = \; = e_{6257}$$

$(300, \quad (10K,1))$
$10K)$

$(300,1)$

$$\rightarrow E \cdot O_j = e_j$$
$$= \text{embedding for word } j$$
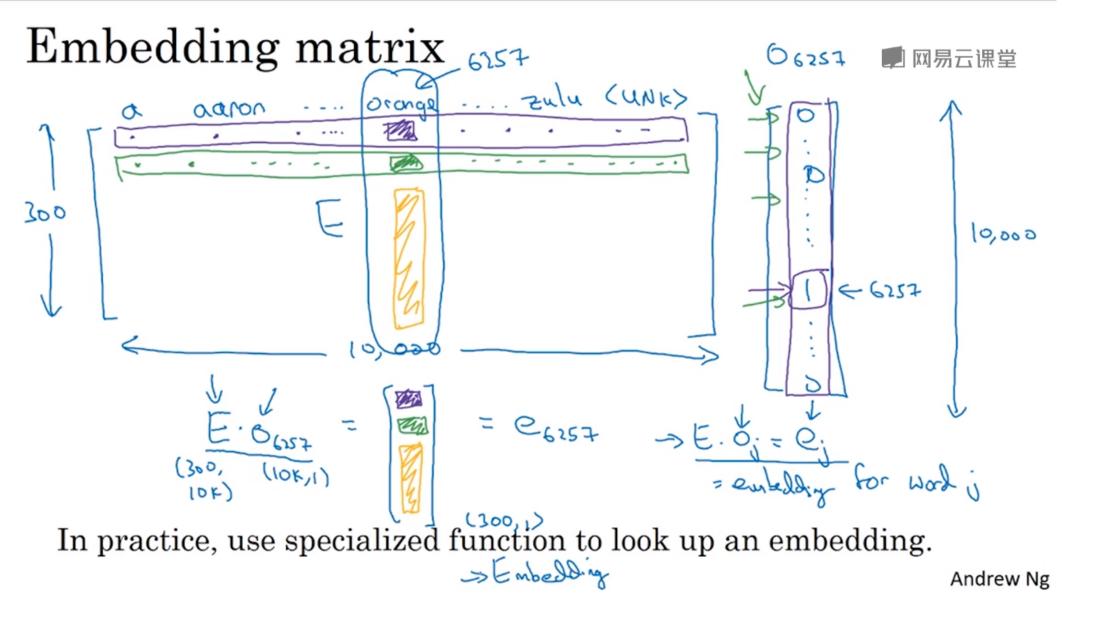
In practice, use specialized function to look up an embedding.

→ Embedding

Andrew Ng

The key terminology is this matrix capital E, which contains all the embeddings for the words of the vocabulary.

So it's actually not efficient to use a matrix vector multiplication to implement this because you will be multiplying a whole bunch things by zero.
And so in practice, you actually use a specialized function to just look up a column of the matrix E rather than do this with the matrix multiplication.