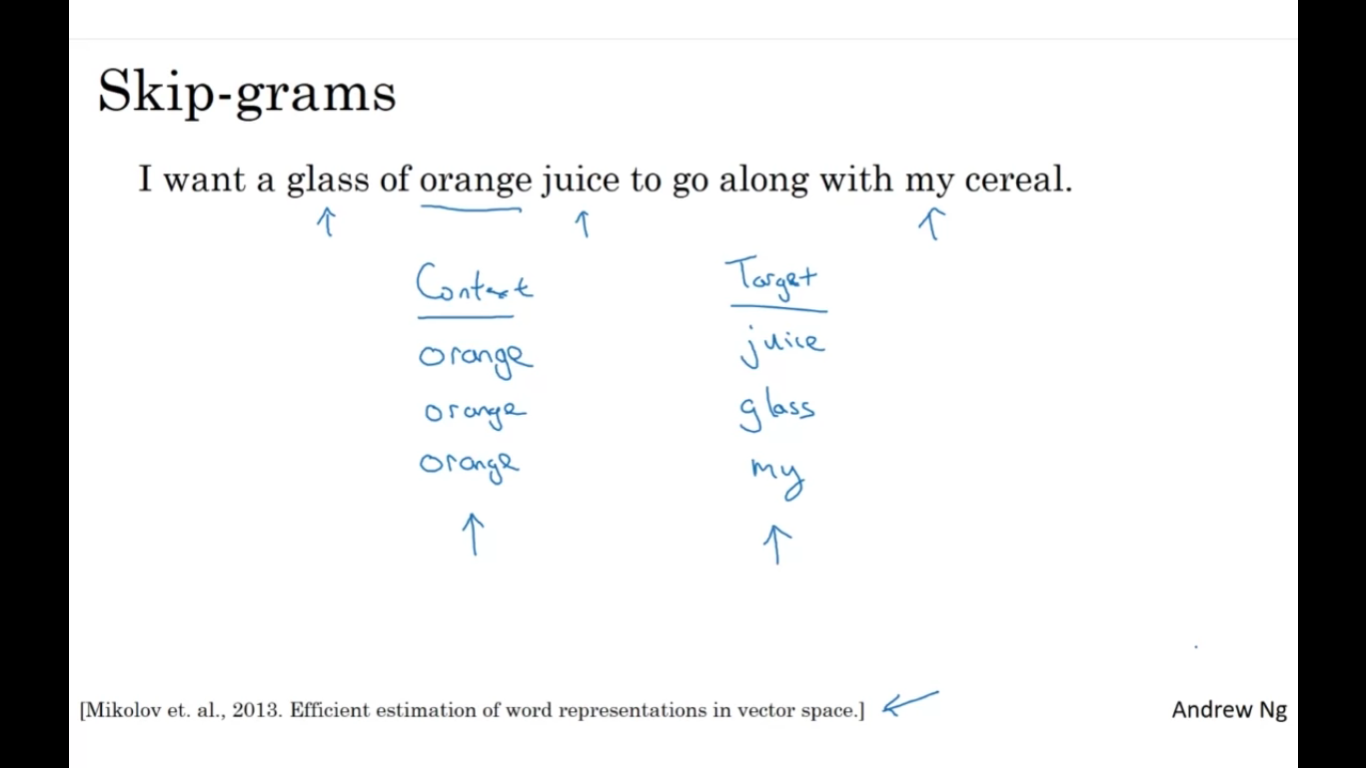
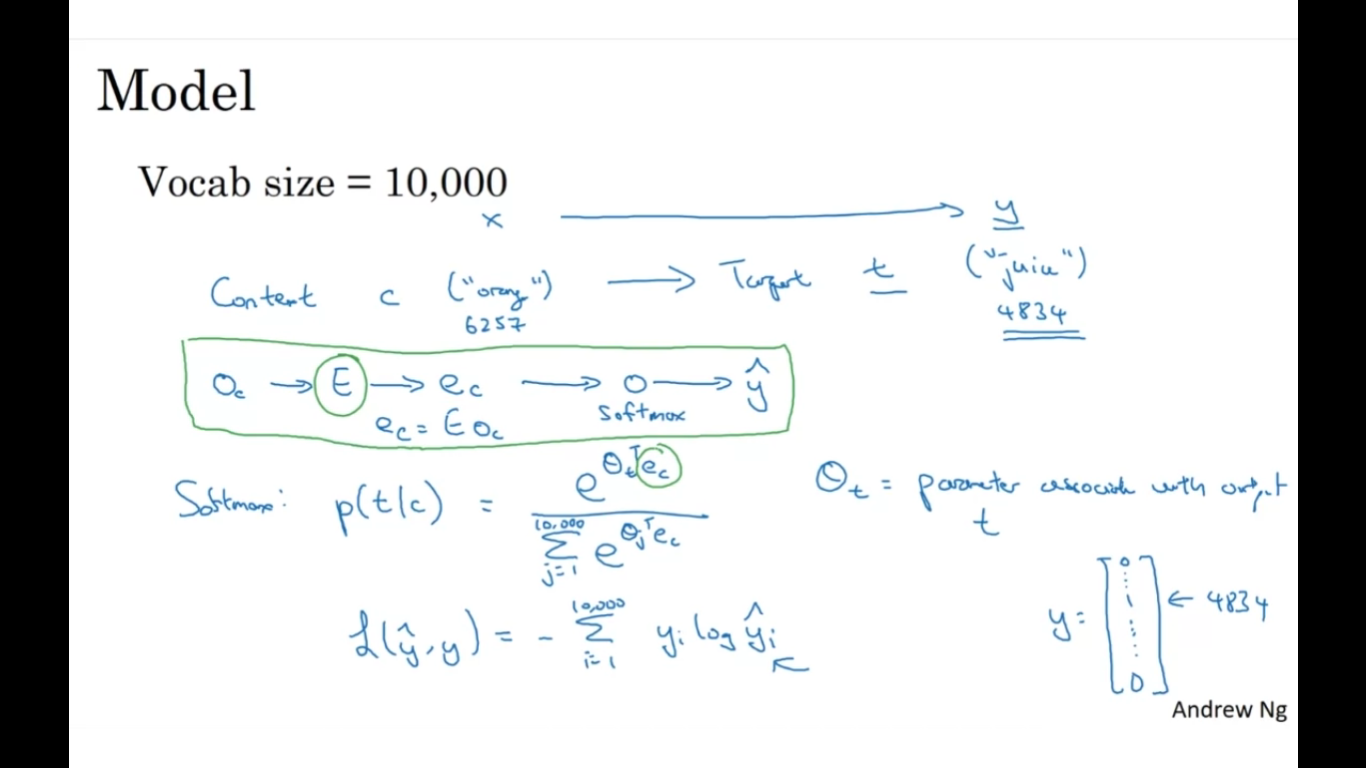
**Word2Vec**

Skip gram model:

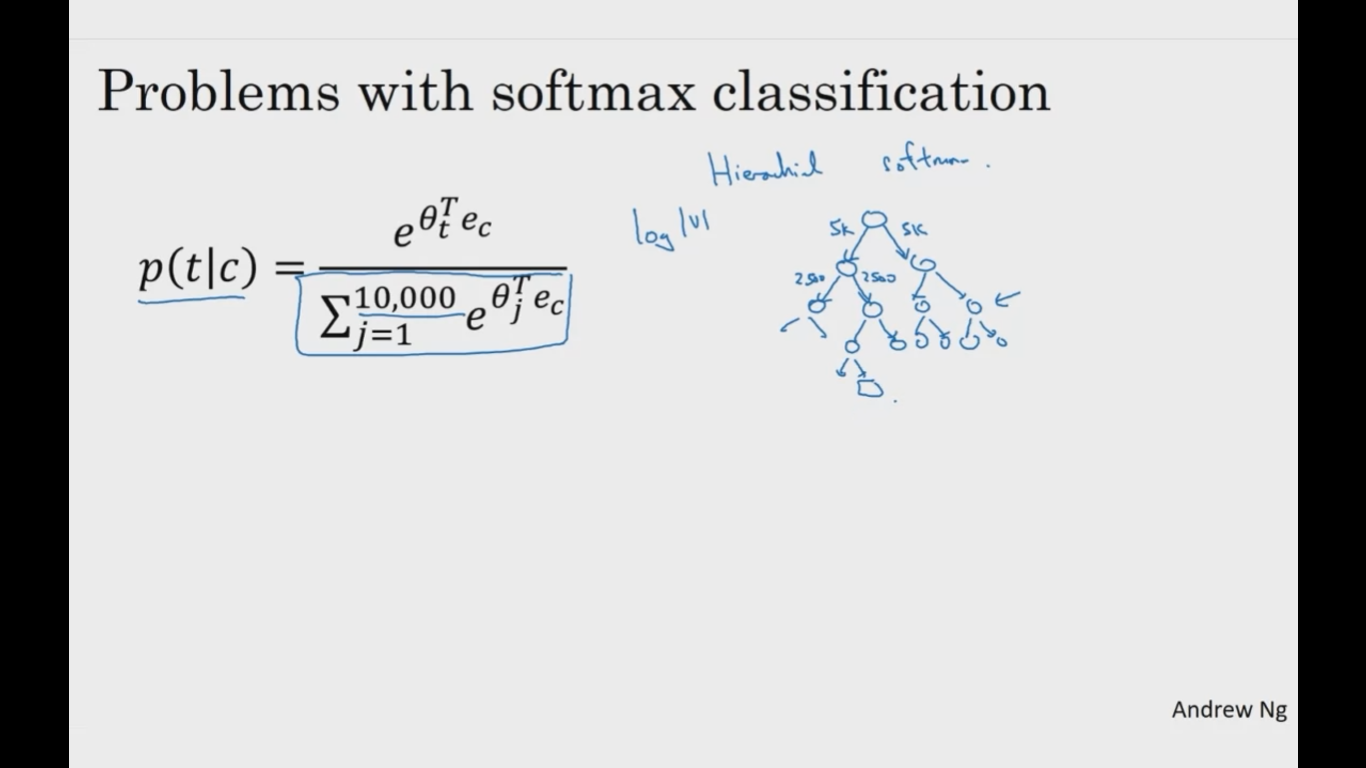
Here we pick randomly the context and the target by choosing a + or – 5 window or so



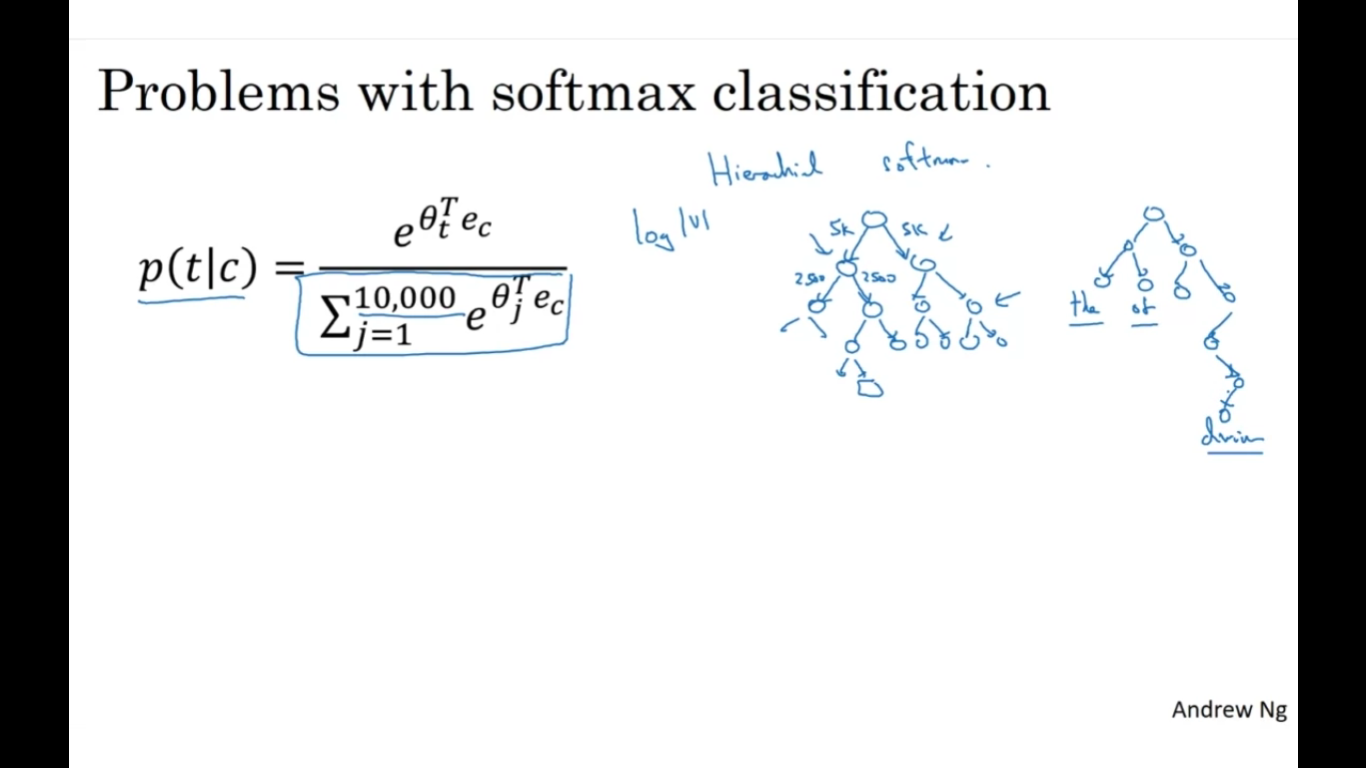


Disadvantages:

Computationally inexpensive (can use hierarchical softmax to overcome)



Hierachical softmax doesnot use balanced tree but it branches its tree using more common words at the top and least common words at the bottom.



how to sample the context C. So once you sample the context C, the target T can be sampled within, say, a plus minus ten word window of the context C, but how do you choose the context C? One thing you could do is just sample uniformly, at random, from your training corpus.

When we do that, you find that there are some words like the, of, a, and, to and so on that appear extremely frequently.

And so, if you do that, you find that in your context to target mapping pairs just get these these types of words extremely frequently, whereas there are other words  like orange, apple, and also durian that don't appear that often.

And maybe you don't want your training site to be dominated by these extremely

frequently or current words, because then you spend almost all the effort updating

ec, for those frequently occurring words.

But you want to make sure that you spend some time updating the embedding, even for

these less common words like e durian.

So in practice the distribution of words pc isn't taken

just entirely uniformly at random for the training set purpose, but

instead there are different heuristics that you could use in order to balance out

something from the common words together with the less common words.

