

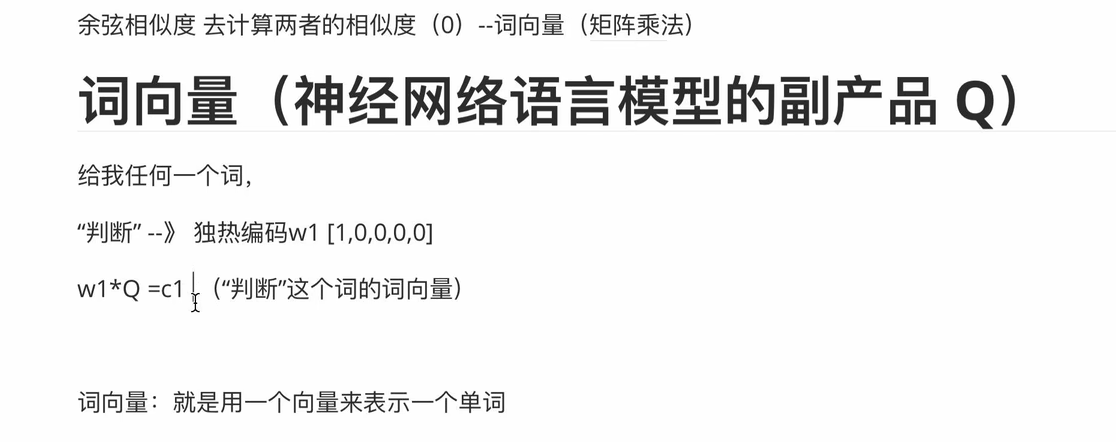
Tanh用于降低计算复杂度

Q可学习，ci是预测的词

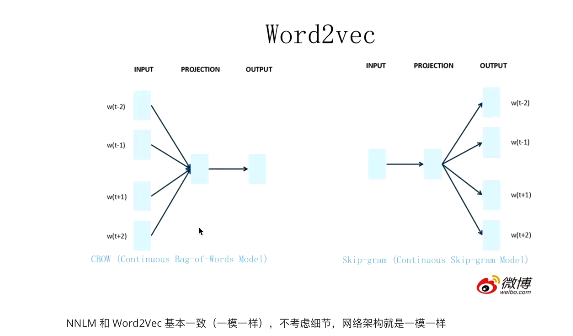
Softmax后会得到一个一维向量，长度为预测后全部词的长度.这里是预测4个词后的词，长度为5

但是独热编码没办法发现词之间的关系

即余弦相似度为0

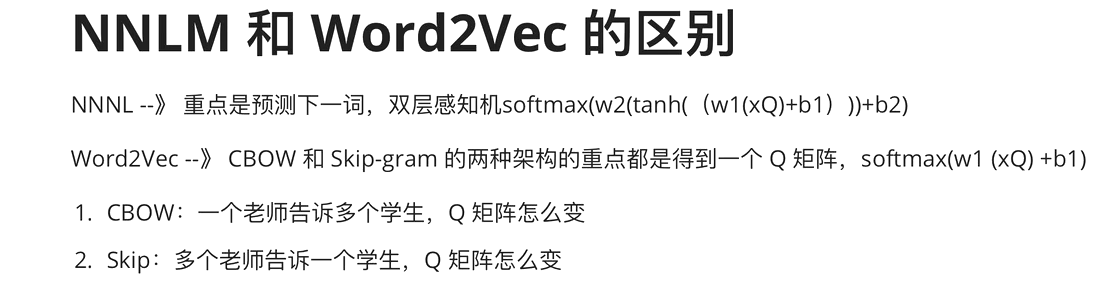


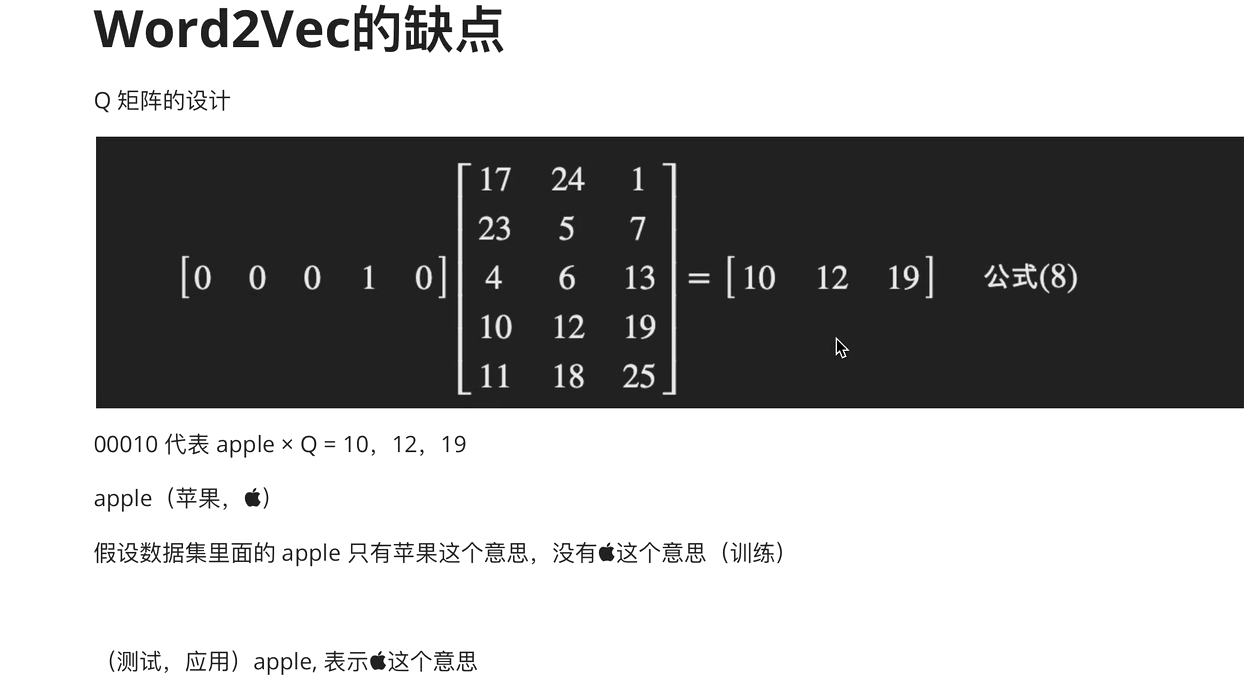
为了得到Q矩阵



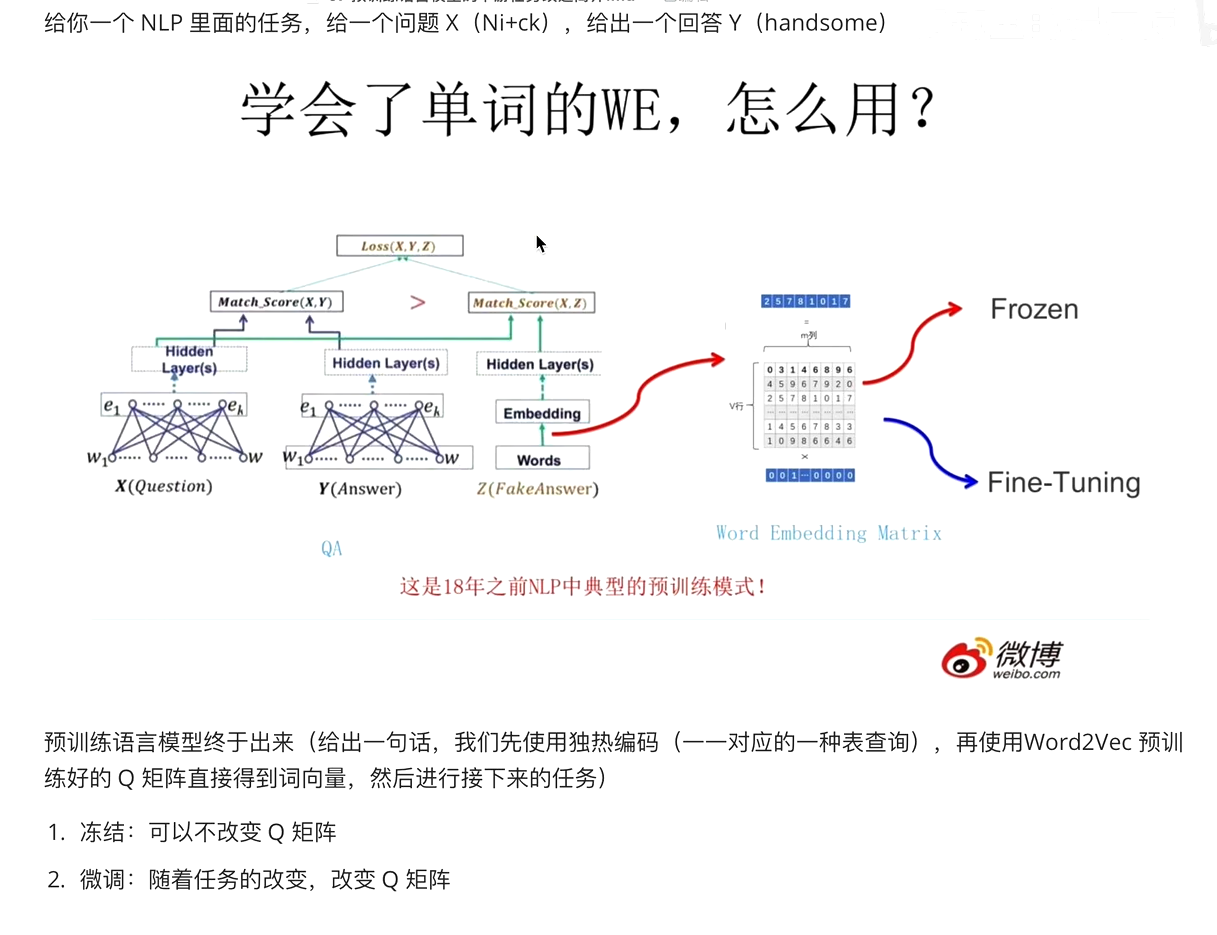
Word2Vec两个小类：CBOW

SKIP





出现歧异



ELMO模型（解决多义词问题）

