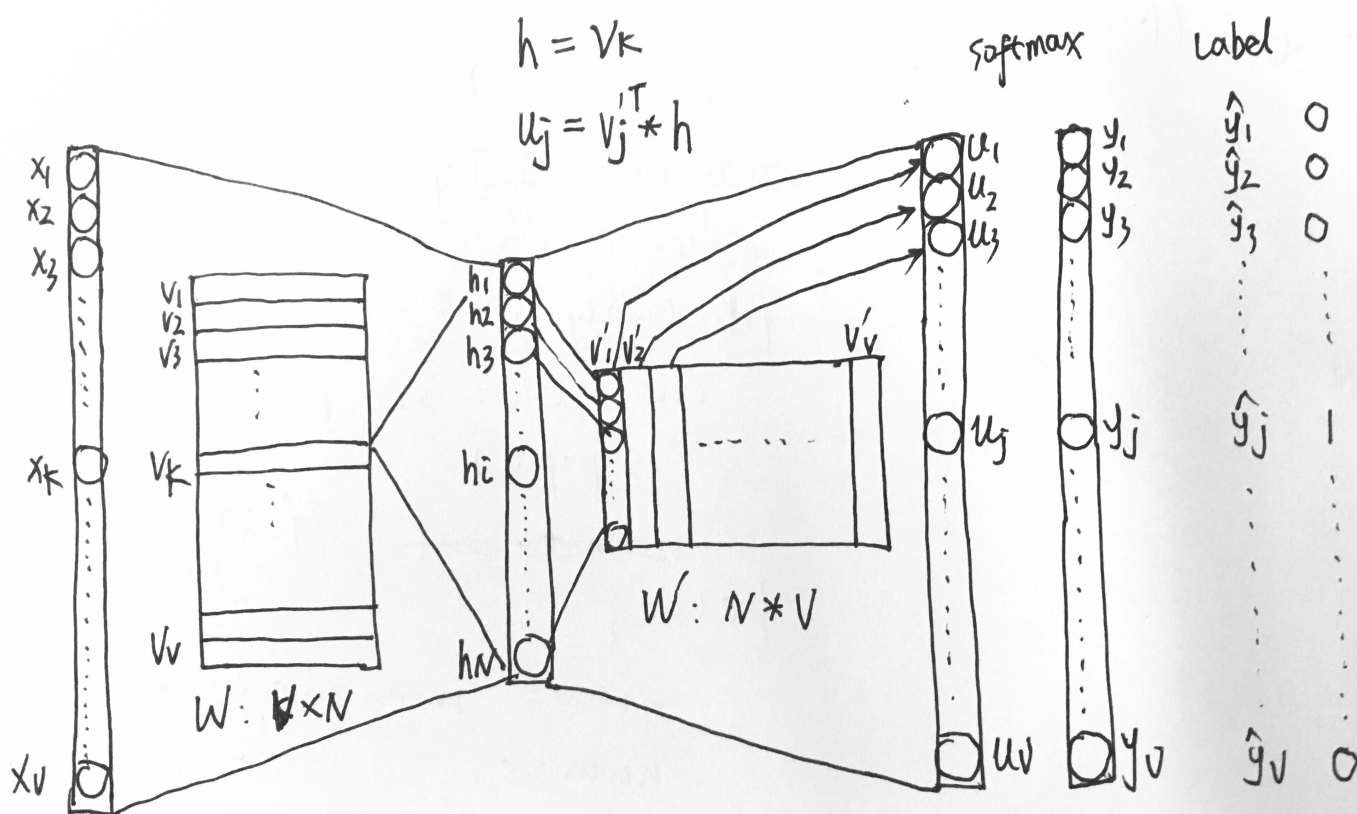


word2vec 数学推导

word2vec 是一种训练词向量的工具，可以将词库中的所有词语映射到 k 维的向量中。主要有两种训练方式，CBOW 和 skip-gram，CBOW 根据上下文来预测目标单词，skip-gram 根据目标单词来预测上下文。下面进行两种方式的数学推导。

CBOW

1. One-word context



假设词表大小为 V ，隐藏层大小为 N ，input-hidden 权重矩阵为 $W_{V \times N}$ ，其中第 k 行的转置为 v_k ，hidden-output 的权重矩阵为 $W'_{N \times V}$ ，其中第 j 列为 v'_j 。计算过程如下：

$$h = v_k$$

$$u_j = v_j'^T * h$$

后验概率为： $p(w_j|w_I) = y_j = \frac{\exp(u_j)}{\sum_{j'=1}^V \exp(u_{j'})}$ ，使用交叉熵损失函数：

$$E = - \sum_{j=1}^V \hat{y}_j \log y_j = - \sum_{j=1}^V \hat{y}_j (u_j - \log(\sum_{j'=1}^V \exp(u_{j'})))$$

E 对 u_j 求导：

$$\frac{\partial E}{\partial u_j} = y_j - \hat{y}_j$$

即预测概率值与真实概率值的差。