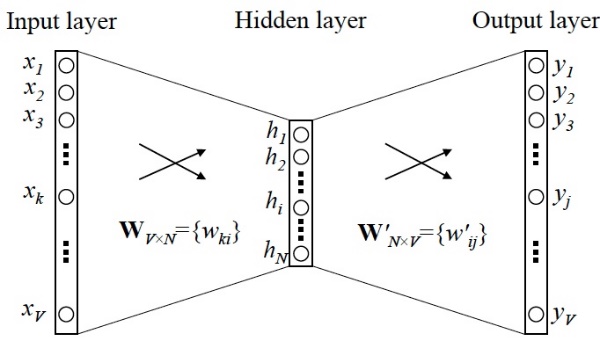
# Word2Vector

论文：Efﬁcient Estimation of Word Representations in Vector Space

通过训练处理简单任务的语言模型，来学习一个合适的词表，将词表示成计算机能够处理的词向量形式，同时这种形式的词向量要能够表示出词与词之间的联系。

常见的有两种形式的语言模型Skip-gram和CBOW，它们都是由输入层、隐藏层、输出层构成。



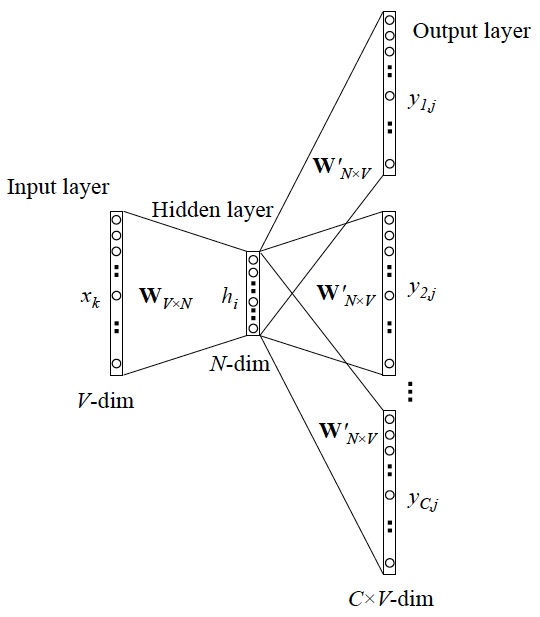
模型最终目的都是获取到输入层和隐藏层之间的权重矩阵，这样对于任意一个词的输入X，都可以得到其词向量。

在模型中，一般有两个词表，一个是中心词表，一个是上下文词表，一个词在作为中心词和上下文词时它的向量表示是不同的。

## Skip-gram

### 模型介绍

模型目标：输入一个词，预测该词的上下文，T是预测窗口的大小。



### 公式推导

变量定义：

：中心词词表

：上下文词词表

：词向量的维度

：词表大小，即词的个数

：中心词

：中心词预测出的上下文单词

：上下文窗口大小

1. 一开始会随机初始化U和V两个词表，shape都是
2. 对于模型而言，其目标就是最大化的概率
3. 1. Softmax函数是用于将一组数据转换成概率分布的一种方式
   2. 使用这种方式就表示出词表中任意一个词是中心词的上下文单词的概率
4. 损失函数，目标是最大化，但一般计算机都适合处理最小化问题，所以可以进行一个转换
5. 对进行开导

模型的目标就是最小化该梯度值，使得达到极值

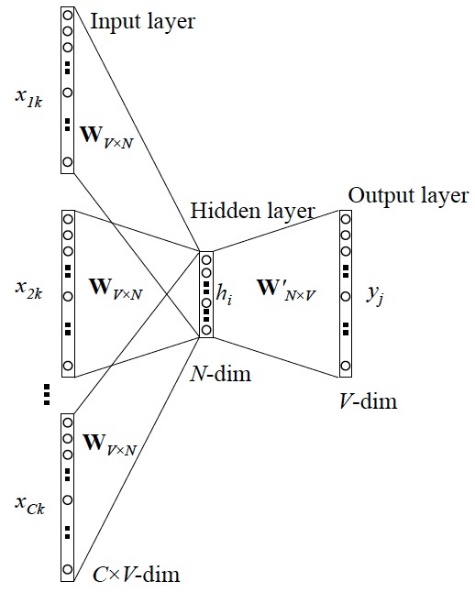
1. 在该模型中，两个词表都是需要优化的参数，所以和都需要更新，都会被当做底进行开导

### 负采样(Negative sample)

1. 其中的分母部分需要计算所有的词，计算量太大，因此采用负采样的方法从词表中选取k个词来计算，而不是所有的词。此时，分母变成
2. 采样的方法基于词频

## CBOW

模型目标：输入词的上下文，预测该词



# Word Embedding

## 首次提出

《A Neural Probabilistic Language Model》

## 思想

## 做法

## 缺点

无法处理同义词问题，一个词最终只能被映射到一个词向量上，但一个词可以有多种不同的含义，使用在不同的场景下。

# ELMO

## 论文

《Deep contextualized word representation》

## 思想

## 做法