# 语言模型

## 训练语言模型的目的

通过语言模型学习词之间的联系，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。

## 语言模型的两种形式

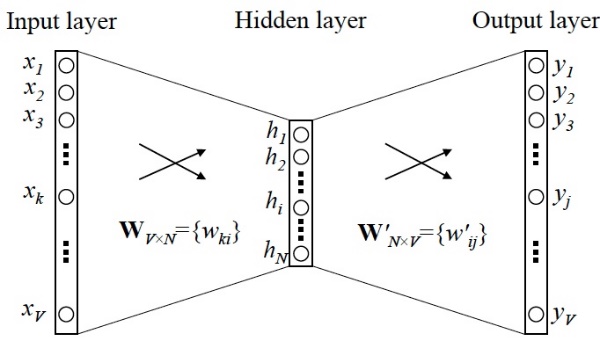
论文：Efﬁcient Estimation of Word Representations in Vector Space

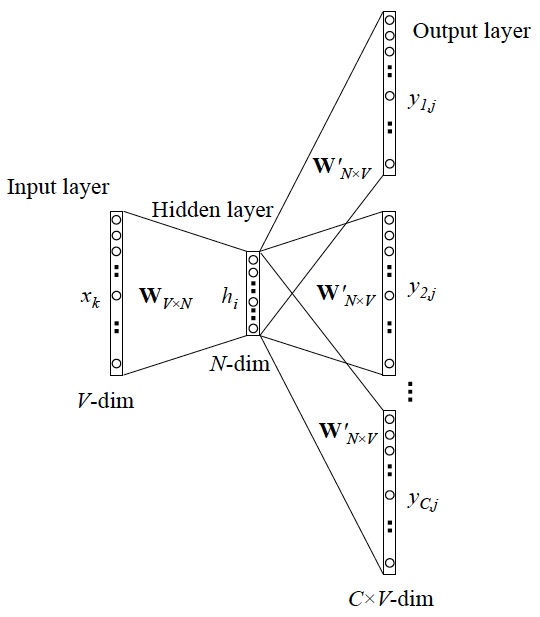
这两种形式的模型最终目的都是获取到输入层和隐藏层之间的权重矩阵，这样对于任意一个词的输入X，都可以得到其词向量。

输入X的表示也有很多种方式：one-hot，随机生成的n维向量

### Skip-gram

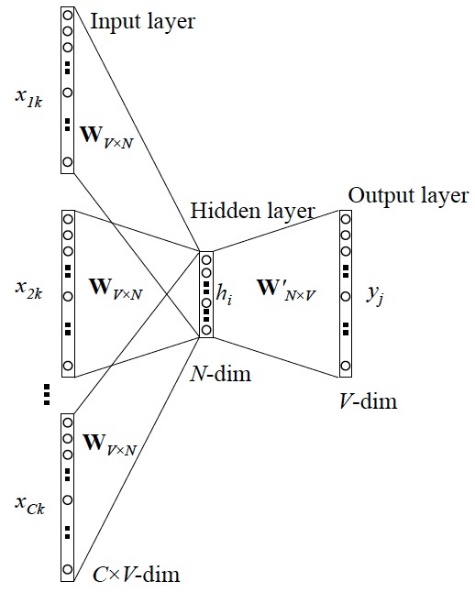
模型目标：输入一个词x，预测该词的上下文y或者{y}(多个上下文)





### CBOW

模型目标：输入词的上下文，预测该词



## 模型改进

论文：Distributed Representations of Words and Phrases and their Compositionality

在Skip-gram的基础上进行了改进

### 以短语为单位

原始的Skip-gram模型是以词为单位，新的模型中以短语为单位，解决习惯用语学习不好的问题。比如““Air Canada”是加拿大航空，而不是简单的Air+Canada。

### Hierarchical Softmax

对Softmax函数的改进，降低了计算开销，引入了随机性，提高了模型的泛华性。

### Negative Sampling

？

### Subsampling of Frequent Words

减少常见词的采样频率，提高非常见词的效果

# 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